



## D4.3 Data Privacy And Security On The Platform

Deliverable ID:	D4.3
Deliverable Title:	Data Privacy And Security On The Platform
Revision #:	1.0
Dissemination Level:	Public
Responsible beneficiary:	CGI
Contributing beneficiaries:	All
Contractual date of delivery:	31.10.2018
Actual submission date:	19.12.2018

# Contents

- Acronyms..... 3**
- 1 Introduction..... 5**
- 2 Security framework..... 5**
  - 2.1 Data types ..... 5**
  - 2.2 GDPR..... 6**
    - 2.2.1 The key concepts ..... 6
    - 2.2.2 The data subject’s rights..... 7
- 3 Optimized city mobility planning..... 7**
  - 3.1 Optimized city mobility planning..... 7**
    - 3.1.3 Data sensitivity ..... 7
    - 3.1.4 Data Collection..... 7
    - 3.1.5 Data Storage and Access ..... 8
- 4 Developing smart HVAC systems that ensure a healthy indoor environment ..... 8**
  - 4.1 Smart HVAC systems that ensure a healthy indoor environment ..... 8**
    - 4.1.1 Data Sensitivity ..... 10
    - 4.1.2 Data Collection..... 12
    - 4.1.3 Data Storage and Access ..... 13
  - 4.2 Intelligent air quality management system ..... 13**
    - 4.2.4 IAQ/OAQ Sensor Data Sensitivity ..... 14
    - 4.2.5 IAQ/OAQ Sensor Data Collection ..... 15
    - 4.2.6 IAQ/OAQ Sensor Data Storage and Access ..... 15
- 5 Promoting independence of specific vulnerable groups ..... 15**
  - 5.1 Rehabilitation decision support ..... 15**
    - 5.1.7 IAQ and OAQ Sensors ..... 16
    - 5.1.8 Patients’ feedback..... 16
    - 5.1.9 Physiological Data..... 17
  - 5.2 Indoor air quality improvement at school..... 17**
    - 5.2.10 IAQ Sensors ..... 17
    - 5.2.11 Self-Report Questionnaires..... 18
    - 5.2.12 Physiological Data ..... 19
    - 5.2.13 3D Motion Camera Data ..... 19
  - 5.3 Tracking of athletes with wearable sensors ..... 20**
- 6 Conclusion ..... 21**

## Acronyms

<b>API</b>	Application Programming Interface
<b>BT</b>	Bluetooth
<b>CO2</b>	Carbondioxide
<b>CRA</b>	Ceske Radiokomunikace company
<b>DPO</b>	Data Protection Officer
<b>EEA</b>	European Economic Area
<b>EU</b>	European Union
<b>GUI</b>	Graphical User Interface
<b>IAM</b>	Identity and Access Management
<b>GDPR</b>	General Data Protection Regulation
<b>HVAC</b>	Heating, Ventilation and Air Conditioning
<b>IAQ</b>	Indoor Air Quality
<b>IoT</b>	Internet of Things
<b>IP</b>	Internet Protocol
<b>JSON</b>	JavaScript Object Notation
<b>JWT</b>	JSON Web Token
<b>LoRaWAN</b>	Long Range Low Power Wide Area Network
<b>M2M</b>	Machine to Machine
<b>MFA</b>	Multi factor Authentication
<b>ML</b>	Machine Learning
<b>MQTT</b>	Message Queuing Telemetry Transport
<b>NGO</b>	Non-governmental Organization
<b>NO2</b>	Nitrogen dioxide
<b>O2</b>	Oxygen
<b>OAQ</b>	Outdoor Air Quality
<b>OS</b>	Operating system
<b>RBAC</b>	Role-based Access Control
<b>POI</b>	Point of Interest
<b>RDS</b>	Amazon Relational Database Service
<b>REST</b>	Representational State Transfer (Type of API)
<b>RF</b>	Radio Frequency
<b>SAS</b>	Shared access signature

<b>SQL</b>	Structured Query Language
<b>SRS</b>	Software Requirements Specification
<b>SRS</b>	Spatial Reference Systems
<b>SSL</b>	Secure Socket Layer
<b>SW</b>	Software
<b>TCP</b>	Transmission Control Protocol
<b>TLS</b>	Transport Layer Security
<b>UI</b>	User Interface
<b>UTM</b>	Urchin Tracking Module
<b>VLCi</b>	Valencia smart City Platform
<b>VM</b>	Virtual Machine
<b>VOC</b>	Volatile Organic Compound

# 1 Introduction

The objective of this deliverable is to provide a more in-depth look of the data security and privacy measures of the different platforms of ESTABLISH. Strong data security is key in creating a functional data platform or user interface and preventing confidential data from falling into the wrong hands. This is why industry-wide data security best practices and guidelines will be applied in securing the data in all of the establish pilots.

ESTABLISH consists of different pilots with different scopes. This report will go over how each of these pilots will apply the security features on their pilot platform. The deliverable is part of work package 4 which is about connecting and managing the sensors. In practice this means collecting the data and managing it in a secure and private way that is compliant with industry standard and local legislation. Thus this deliverable is part of task 4.3 which is the final task of the package. D4.3 will be followed by the final deliverable or D4.4. where all the previous deliverables (D4.1 data acquisition, D4.2 data management and D4.3 data security and privacy) will be implemented.

The deliverable is formatted so that first the security framework is presented. Then each pilot discusses their data privacy and security issues. Each collected data type is discussed in terms of sensitivity, collection as well as storage and access. These chapters give an idea on what is done in the pilot in terms of those processes and how are security and privacy taken into account. Data sensitivity chapters give an idea on how sensitive is the collected data in nature thus guiding the collection, storage and access processes.

## 2 Security framework

### 2.1 Data types

When referring to data protection in this deliverable we are considering the following groups of data:

- Personal data
- Self-report data
- Environmental data collected by sensors, outdoor and indoor as well as motion data
- Psychological data collected by wearables, and
- Open data.

The source of data is different for each type of data and so are the measures to protect it and keep it private. Below you can see a summary in table 1 of the collected data in the ESTABLISH project. The data types are in order from most sensitive to least sensitive from the left to right.

*Table 1. Summary of collected data types in ESTABLISH*

Pilots	Personal Data	Psychological data	Environmental data	Self-report data	Open data
Optimized City mobility planning					X
Smart HVAC systems			X		

Intelligent air quality management			X		
Rehabilitation decision support	X	X	X	X	
Indoor Air Quality improvement at schools		X	X	X	
Tracking athletes with wearables	X	X	X		

The first two in the table 1 are considered personal data while the rest are, as such, non-identifiable to any individual. However, special care must be taken if the non-personal data can be combined in a way that it is possible to identify a person by combining the two pieces of data. Both personal data management practices and the risks of non-personal data being combined to form an identifiable piece of information, are discussed in the following chapters.

## 2.2 GDPR

The General Data Protection Regulation (GDPR) is a European Union (EU) regulation regarding data protection and privacy for people in the European Union and European Economic Area (EEA). Additionally it concerns exporting data outside EU and EEA. In short, this means that if any party of the data management process – data controller, data processor or data subject is based in the EU or EEA areas, the GDPR will have to be complied with in the process.

GDPR categorizes the parties of the process into these three different groups - data controllers, data processors and data subjects. A data controller is the organization or individual that determines the purposes and means of the processing of personal data. A data processor is the party that processes the data on behalf of the data controller. Finally, the data subject is individual, whose data is being collected and processed.

While the data processor can be held liable only for breaching the data controller’s instruction or not complying with the GDPR processor-specific obligations, the controller is liable for general non-compliance of the GDPR.

The content of the GDPR can be grouped into two – the key concepts and the data subject’s rights: the key concepts and the data subject’s rights (presented below).

### 2.2.1 The key concepts

**Purpose limitation** – Personal data should be collected and processed for predefined tasks only and cannot be repurposed without additional consent from the data subject.

**Data minimization** – Only the minimum amount of data required to accomplish the defined task should be used in the data processing.

**Data accuracy** – Data controllers should make sure that collected personal data is accurate and kept up to date. “Every reasonable step must be taken” to erase or rectify inaccurate or incomplete data.

**Storage limitation** – Data controllers must not retain personal data for longer that necessary.

**Confidentiality and Integrity** – The data controllers must ensure that the personal data is “processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or

unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures”.

**Lawfulness and Transparency** – Data controllers must make sure that their data collecting practices comply with the law and that they are not hiding anything from the data subject. The reasons why certain data is collected should be stated in the data controllers’ privacy policy.

### 2.2.2 The data subject’s rights

**Breach Notification** - The data subject will have to be notified of a data breach within 72 hours unless the breach is unlikely to result in a “risk for the rights and freedoms of individuals”. Additionally the data processor will have to notify the data controller immediately after becoming aware of a data breach.

**Right to Access** - The data subject has the right to query the data controller for a confirmation whether their personal data is being processed, where and why. The data controller is also obligated to provide an electronic copy of the data free of charge if the data subject requests so.

**Right to Erasure** – The data subject has the right to have the data controller erase their personal data, suspend the further spreading of this data and possibly have the data processor stop processing the data. The right can be used if one of the following ground applies – the data is no longer relevant to the original purposes for the processing, the data subject withdraws consent to process the data, the personal data has been unlawfully processed or the personal data has to be erased because of a legal obligation in Union or Member State law.

**Data Protection By Design & By Default** – The data controllers are obligated to design their business processes and practices so that the data protection is built into them.

**Data Protection Officers** – Data Protection Officers (DPO) must be “appointed for all public authorities, and where the core activities of the controller or the processor involve ‘regular and systematic monitoring of data subjects on a large scale’ or where the entity conducts large-scale processing of ‘special categories of personal data’”.

## 3 Optimized city mobility planning

### 3.1 Optimized city mobility planning

The pilot Optimized City and Mobility Planning will build an advanced application for providing planning services and mobility information both for citizens and for city authorities considering relevant information such as contamination or traffic conditions.

#### 3.1.3 Data sensitivity

In this pilot, all the data used comes from open data platforms or third-party web services. No sensors are provided by ESTABLISH partners.

The pilot is being developed in Valencia (Spain) and the data comes from the Valencia smart City Platform (VLCi) which is the name of its smart city platform. These data sources are freely available for anybody, without copyright, patent or any other restriction

In addition to the data obtained from the open data, the use case will make use of public weather forecast systems; this information is relevant to make predictions of pollution levels.

#### 3.1.4 Data Collection

The use case will provide a mobile application to calculate routes and send notifications to users. In no case the system will make use of personal information such as name or address that allows users to be

identified. The only requested information to calculate the route is the origin, destination and the preferences of the user (preferred means of transport)

### 3.1.5 Data Storage and Access

The data base used to store the information retrieved from open data sources, generated by several analysis modules (prediction, route planner) and mobile application will be ElasticSearch. Elasticsearch is a search engine based on Lucene. It provides a distributed, multitenant-capable full-text search engine with an HTTP web interface and schema-free JSON documents. Elasticsearch can be used to search all kinds of documents. It provides scalable search, has near real-time search, and supports multitenancy. Elasticsearch makes all its features available through the JSON and Java API.

Additionally, a PostgreSQL database is used to store the user personal information, such as the name, the email and the address, the user preferences and the vehicle's main feature. In order to protect all data stored, the security of PostgreSQL database server will be hardened as follows:

- **Client Authentication Control:** It is possible to define a configuration file as a set of records, one per line, where each record specifies a connection type, a client IP address range (if relevant for the connection type), a database name, a user name, and the authentication method to be used for connections matching these parameters. The first record with a matching connection type, client address, requested database, and user name is used to perform authentication.
- **Server Configuration:** It is possible to enhance the security through a configuration file by controlling which ips will be allowed to connect to the server.
- **Data Encryption:** PostgreSQL has native support for using SSL connections to encrypt client/server communications for increased security. SSL (Secure Sockets Layer) is the standard security technology for establishing an encrypted link between a web server and a browser. This link ensures that all data passed between the web server and browsers remain private and integral.

Other security measures to be considered if it is necessary are as follows:

- **User and Role Management,** that is, grant users the minimum amount of access they need.
- **Logging,** that is, Postgresql provides a wide variety of configuration parameters for controlling what, when, and where to log.
- **Auditing,** that is, The PostgreSQL Audit Extension (pgAudit) provides detailed session and/or object audit logging via the standard PostgreSQL logging facility.

## 4 Developing smart HVAC systems that ensure a healthy indoor environment

### 4.1 Smart HVAC systems that ensure a healthy indoor environment

In the Smart HVAC systems that ensure a healthy indoor environment-pilot, we have been developing the HVAC system that autonomously learn behavior patterns of the users/inhabitants of the building and take advantage of this knowledge to get the building ready for the predicted needs. It will try to resolve



the tension between energy efficiency and quality of indoor climates that occurs e.g. after retrofitting existing buildings by offering an affordable solution.

We previously marked the key factors of and benefits of LoRaWAN technology one of which is the layer of security for the network and one for the application with AES-128 key-encryption.

The AES encryption algorithm defines a number of transformations that are to be performed on data stored in an array. The first step of the cipher is to put the data into an array; after which the cipher transformations are repeated over a number of encryption rounds. The key length with 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256-bit keys, determines the number of rounds. The message structure is presented in figure 1.

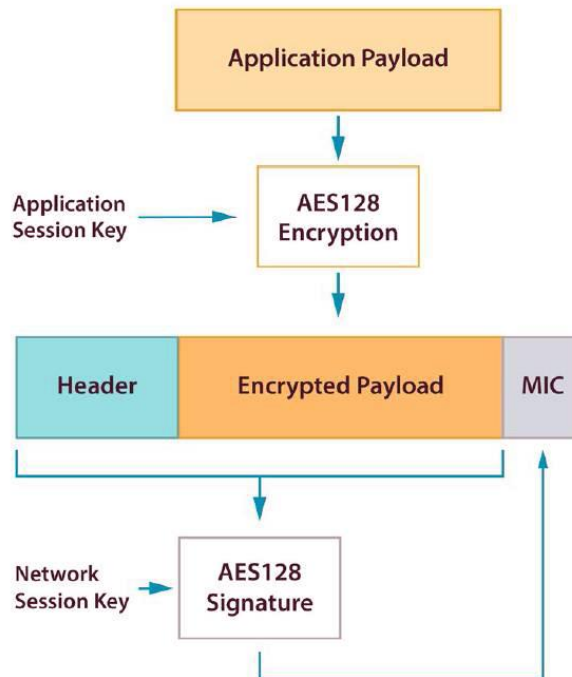


Figure 1. Message structure

- The network key eliminates fake devices in the network, duplicating and diverting messages.
- The application key ensures privacy of the entire infrastructure to the user.
- The transmitted quantity is encoded into the payload of the message.

Public networks such as the Internet do not provide a means of secure communication between entities. Communication over such networks is susceptible to being read or even modified by unauthorized third parties. Cryptography helps protect data from being viewed, provides ways to detect whether data has been modified, and helps provide a secure means of communication over otherwise no secure channels. Data encrypted by using a cryptographic algorithm, transmitted in an encrypted state, and later decrypted by the intended party. The range of encryption is illustrated in figure 2. If a third party intercepts the encrypted data, it will be difficult to decipher. In the .NET Framework, the classes in the System.Security.Cryptography namespace manage many details of cryptography for you. Some are wrappers for the unmanaged Microsoft Cryptography API (CryptoAPI), while others are purely managed implementations. You do not need to be an expert in cryptography to use these classes. When you create a new instance of one of the encryption algorithm classes, keys are auto generated for ease of use, and default properties are as safe and secure as possible.

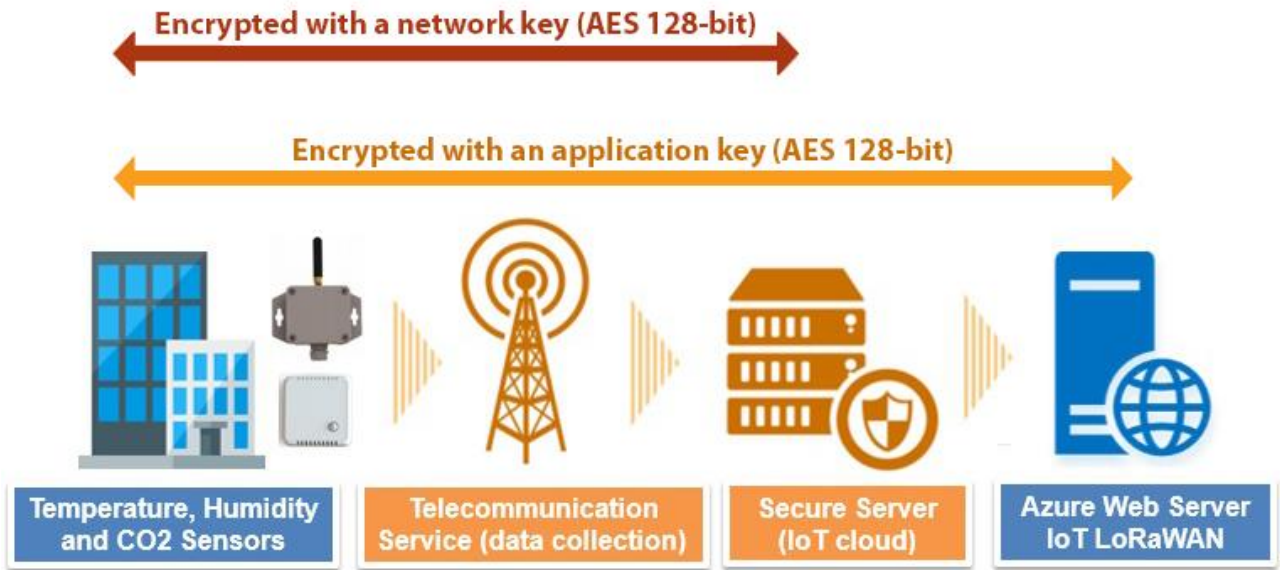


Figure 2. Range of encryption




The personal data of the client such as the full name, login and password for entering the system is encrypted, they are securely stored in a central database and can be accessed only by authorized persons.

IMA will follow necessary privacy principles during both development and the subsequent process. Privacy concerns of the end users are taken into consideration during every step of the process. We will try to identify possible risks in advance and to ensure, that adequate solutions are put in place.




#### 4.1.1 Data Sensitivity

The sensors specified in the table 1 are usually installed in a home, a school or an office building. The data collected from the devices is composed of device identifier and raw sensor data which does not contain any private data such as location information of the home or the office building.

Table 2. IAQ and OAQ Sensors

RisingHF / RHF1S001	Ascoel / CO868LR - COUS915LR	Elsys / ERS
		
Temperature, Humidity	CO2, Temperature and Humidity	CO2, Humidity, Light, Motion and Temperature

<ul style="list-style-type: none"> <li>• Powered by Lithium-thionyl chloride battery</li> <li>• 5 years of operation for 1 uplink per minute.</li> <li>• Extended industrial operating temperature: -40°C to +85°C.</li> <li>• Outdoor use: IP64 enclosure.</li> <li>• Accuracy: ±5% RH typically from 20% RH to 80% RH at 25°C. ±0.5°C typically from +5°C to +60°C.</li> <li>• LoRa WAN compatible: Class A, uplink rate programmable from 1 minute to 24 hours.</li> <li>• Operating frequency bands (Option at order): 434MHz, &lt;10mW radiated power. 868MHz, &lt;25mW radiated power. 920MHz, &lt;25mW radiated power.</li> </ul>	<ul style="list-style-type: none"> <li>• Dual wave length NDIR (non dispersive infrared technology)</li> <li>• Accuracy at 25°C and 1013mbar of 5000ppm</li> <li>• Temperature range -10°C &lt; T &lt; +55° C with typical accuracy tolerance +/- 0.3°C</li> <li>• Humidity range 0 &lt; RH &lt; 100%</li> <li>• LoRa TM long range 868MHz radio module</li> <li>• LoRAWAN v1.0 class A compliant</li> <li>• Programmable Alive signal</li> <li>• One or two thresholds for each CO2, T and RH programmable from remote</li> <li>• CO2 timing measurement programmable from remote</li> <li>• Graphic LCD resolution 128x64</li> <li>• Up to 5 five messages of 21 character each can be displayed on 5 rows on the LCD</li> <li>• Buzzer for low battery indication</li> <li>• Buzzer for acoustic signaling</li> <li>• Lithium-thionyl 6.0 Ah, C size.</li> </ul>	<ul style="list-style-type: none"> <li>• LoRa Alliance Certified</li> <li>• Temperature</li> <li>• Humidity</li> <li>• Light</li> <li>• Motion (PIR)</li> <li>• CO2 (ERS-CO2)</li> <li>• Sound, peak and average (ERS-sound)</li> <li>• NFC for easy configuration</li> <li>• Size : 86x86x26mm</li> <li>• Accuracy: ± 0.5°C, ±2%rh</li> <li>• Resolution: 0.1°C, 0.1%rh</li> <li>• Approx. range: 8km*</li> <li>• Battery life: 10 years**/Approx. 3 years for ERS-sound</li> <li>• US902-928, EU863-870, AS923, AU915-928, KR920-923</li> <li>• 2 x 3.6V AA lithium battery</li> </ul>
--	---	--

Elsys / ELT-1	U-Blox America Inc. / SARA-U201-03B	Protronix / NLII-iVOC+RH+T
		
Accelerometer, Humidity and Temperature	Temperature	VOC, Humidity and Temperature

<ul style="list-style-type: none"> <li>• Analog 0-10V</li> <li>• Digital</li> <li>• Pulse counter</li> <li>• Direct connection to 1-wire temperature sensor</li> <li>• Direct connection to Decagon moisture sensor</li> <li>• Direct connection to S0 outputs</li> <li>• Direct connection to Maxbotix ultrasonic distance sensor</li> <li>• Water leak sensor</li> <li>• Embedded sensors for temperature, humidity and acceleration.</li> <li>• NFC for easy configuration</li> <li>• +14dBm maximum power</li> </ul> <p>The ELT-1 is a 1-channel GPIO wireless transceiver for LoRaWAN™. The module can provide power to external sensors. The power-on time before data sampling can be adjusted.</p> <p>Battery: 3.6V lithium AA-size Approx. range: ±8km @ LoRa™ modulation SF10 Battery life: Up to 10 years, depending on transmission interval.</p>	<ul style="list-style-type: none"> <li>• Ultra low power consumption delivering 10+ years battery life</li> <li>• Excellent extended range in buildings, underground (MCL 164 dB)</li> <li>• Extended temperature range of -40 to +85° C and ISO/TS16949 manufacturing</li> <li>• Easy migration between u-blox 2G, 3G, and 4G modules</li> <li>• Very small SARA LGA form factor for easy manufacturing</li> </ul>	<ul style="list-style-type: none"> <li>• Voltage supply range 14V - 40V DC; 18V -30 V AC</li> <li>• Average consumption 0,5 W</li> <li>• Volatile Organic Compounds – Acetone, Ethanol, Isoprene, Butane measuring range 450 - 2000 ppm</li> <li>• RH measuring range 0 - 100%</li> <li>• RH measurement accuracy ± 3%</li> <li>• Temperature measuring range 0 - 40 °C</li> <li>• Temp. measurement accuracy ±0,4 °C</li> <li>• Operating temperature 0 +50 °C</li> <li>• Storage temperature -20 +50 °C</li> <li>• Life: Up to 10 years, depending on transmission interval.</li> </ul>
---	---	---

Because we're working with a radio protocol, anyone will be able to capture and store messages. It's not possible to read these messages without the AppSKey, because they're encrypted. Nor is it possible to tamper with them without the NwkSKey, because this will make the MIC check fail. It is however possible to re-transmit the messages. These so-called replay attacks can be detected and blocked using frame counters.

When a device is activated, these frame counters (FCntUp and FCntDown) are both set to 0. Every time the device transmits an uplink message, the FCntUp is incremented and every time the network sends a downlink message, the FCntDown is incremented. If either the device or the network receives a message with a frame counter that is lower than the last one, the message is ignored.

This security measure has consequences for development devices, which often are statically activated (ABP). When you do this, you should realize that these frame counters reset to 0 every time the device restarts (when you flash the firmware or when you unplug it). As a result, The Things Network will block all messages from the device until the FCntUp becomes higher than the previous FCntUp. Therefore, you should re-register your device in the backend every time you reset it.

#### 4.1.2 Data Collection

We selected LoRa communication standard as promised technology for potential customers and signal coverage in Czech operated by Ceske Radiokomunikace company (CRA).

The network is built on LoRa technology, which allows wireless connection of intelligent devices at great distances with minimum energy requirements. This is why it is ideal for connecting sensors, detectors, meters and other devices in the field to our system. End sensors provide data (data collection) at Telecommunication service provided by CRA transferred via dedicated secure network to Secure Server (IoT Cloud).

The advantages of CRA and the standard LoRa sensor include:

- Secure communication from the device
- Infrastructure located in the Czech Republic
- Secure data repository Coded data transfer

Chapter 4.1 describes methods for encrypting data with AES-128 key-encryption during transmission from the end device to CRA Telecommunication service (data collection) and to our server.

### 4.1.3 Data Storage and Access

After receiving information from the sensors and processing it, the data is stored in special tables. Azure SQL Database is the fully managed cloud equivalent of the on-premises SQL Server product that has been around for decades. We use store relational data in a transactional manner with advanced querying capabilities, Azure SQL Database is the best service for us.

Azure SQL Database secures your data by limiting access to your database using firewall rules, authentication mechanisms requiring users to prove their identity, and authorization to data through role-based memberships and permissions, as well as through row-level security and dynamic data masking.

Azure SQL Database enforces encryption (SSL/TLS) at all times for all connections, which ensures all data is encrypted "in transit" between the database and the client. This will happen irrespective of the setting of Encrypt or TrustServerCertificate in the connection string.

Managing databases and logical servers within Azure is controlled by your portal user account's role assignments. To help protect your data, firewalls prevent all access to your database server until you specify which computers have permission using firewall rules. The firewall grants access to databases based on the originating IP address of each request.

**SQL Authentication method** uses a username and password. For example when you created the logical server for your database, you specified a "server admin" login with a username and password. Using these credentials, you can authenticate to any database on that server as the database owner.

**Azure Active Directory Authentication method** uses identities managed by Azure Active Directory and is supported for managed and integrated domains. Use Active Directory authentication (integrated security) whenever possible. If you want to use Azure Active Directory Authentication, you must create another server admin called the "Azure AD admin," which is allowed to administer Azure AD users and groups. This admin can also perform all operations that a regular server admin can.

In addition, the IoTLoRaWan Server provides data access service with JSON query format describing various query constraints over HTTPS. The server provides user-based authentication mechanism to assure that only valid users can access the storage. The server resides in Azure cloud platform and exploits Azure security services for enhancing data security and privacy such as multi-factor authentication (MFA), role-based access control (RBAC) and data encryption.

## 4.2 Intelligent air quality management system

The Korean pilot aims to develop intelligent air quality management system. The purpose of the system is to provide a healthy indoor air quality in buildings through the adaptive control of air purifiers based on the analysis results of indoor environmental data. The system uses both IAQ (Indoor Air Quality) and OAQ (Outdoor Air Quality) sensor devices to collect environmental data related to the indoor air quality since the indoor air quality is affected by outdoor pollutants as well as indoor pollutants. In addition, it tries to control air purifiers adaptively to enhance indoor air quality. The data management server manages and analyzes the environmental data collected from the IAQ and OAQ devices. This pilot








consists of the following four basic goals in order to keep indoor environment healthy through environmental data analysis.

- IAQ/OAQ sensor devices to measure environmental data
- Collection and management of environmental data
- Analysis of environmental data to extract indoor air quality index
- Adaptive control of air purifiers.

#### 4.2.4 IAQ/OAQ Sensor Data Sensitivity

There is one source of data from IAQ/OAQ devices used in the pilot (table 3). IAQ and OAQ devices integrate many kinds of sensors to measure the various environmental data into one hardware module. The devices also include wireless communication to transmit the measured sensor data. Each IAQ sensor device for indoor air quality measurement includes temperature, humidity, CO<sub>2</sub>, illuminance, noise, VOC (volatile organic compounds), Formaldehyde and dust sensors.

Table 3. IAQ sensor specification

Type	CO <sub>2</sub>	illuminance	Noise	VOC	Formaldehyde	Dust	Humidity/ Temperature
Product Picture							
Measurement Range	0~2000 ppm	10~1,000 Lx	-42dB	1~10ppm	0~500ppb	0~500ug/m <sup>3</sup>	-40~125°C 0~100%RH
Accuracy	<±50ppm+ 2% of measuring value	-	±3dB	Coway's Algorithm	<±30%@Full Range	±15%	±0.3°C ±2%
Interface	I2C	Analogue	Analogue	Analogue	I2C	UART	I2C

An OAQ sensor device for outdoor air quality measurement uses solar energy as its battery charging and is designed to endure the tough outdoor weather (table 4). It includes temperature, humidity and dust sensors.

Table 4. OAQ sensor specification

Type	Solar Cell	Battery	Dust Sensor	Temperature/Humidity	RF Module
Product Picture					
Spec	9V, 3.3W	Li-ion (3.7V, 14Ah)	Laser Type	40~125°C 0~100%RH	447MHz
Interface	Analogue	Analogue	UART	I2C	UART

The IAQ/OAQ devices are usually installed in a home or an office building. The data collected from the devices is composed of device identifier and raw sensor data which does not contain any private data such as location information of the home or the office building. The information where the devices are installed is managed by the backend data management server in which authorized users can access the data.

#### 4.2.5 IAQ/OAQ Sensor Data Collection

An IAQ device communicates with its subordinate OAQ device through Radio Frequency (RF) to gather both IAQ and OAQ environmental data and sends the data to the data management server via Transmission Control Protocol (TCP). The server communicates with the IAQ devices via TCP in order to collect and store the data in PostgreSQL DBMS storage. In order to store the sensor data in the storage, each IAQ/OAQ should be registered in the server. The sensor data coming from unregistered devices is rejected. Every sensor data packet from a IAQ/OAQ device has checksum for checking packet validity. The server checks that the checksum of the packet is valid for storing the data.

#### 4.2.6 IAQ/OAQ Sensor Data Storage and Access

The data management server provides data access service with JSON query format describing various query constraints over HTTP. The server uses Spring framework over Apache Tomcat for HTTP processing and PostgreSQL for environmental data storage. The server provides user-based authentication mechanism to assure that only valid users can access the storage. The server can reside in Azure cloud platform, if needed, in which case the server can exploit Azure security services for enhancing data security and privacy such as multi-factor authentication (MFA), role-based access control (RBAC) and data encryption.

## 5 Promoting independence of specific vulnerable groups

### 5.1 Rehabilitation decision support

The Rehabilitation decision support pilot will combine environmental sensor data with physiological and behavioral sensor data to empower patients in a rehabilitation clinic with decision support tools for behavioral choices and treatment options.

In order to find links between the biological data and environmental conditions it is required to monitor the physical activities (monitoring heart rate, the burned calories, sleep patterns) during the recovery programmes as recommended by the kynetotherapists, trainers, or physical education teacher.

The Rehabilitation decision support pilot will run in the premises of a Romanian NGO, in Bucharest. The decision support system advanced within the Romania pilot is based on the and environmental data analysis.

- IAQ/OAQ sensor devices to measure environmental data
- Collection and management of environmental data
- Analysis of environmental data to extract air quality index
- Wearable devices in order to collect physiological data
- Analysis through machine learning resources both environmental and physiological measurements.

### 5.1.7 IAQ and OAQ Sensors

IAQ sensor data consists of monitoring the following parameters: temperature, relative humidity, atmospheric pressure, pressure difference (in vs out), CO<sub>2</sub>, NO<sub>2</sub>, VOC, and O<sub>2</sub>.

OAQ sensor data consists of monitoring the following parameters: temperature, relative humidity, atmospheric pressure, pressure difference (in vs out), CO<sub>2</sub>, NO<sub>2</sub>, PMs, and O<sub>3</sub>, solar radiation, the level of precipitation, wind and speed direction. The data is collected from 2 different types of sensors.

#### *Air Quality Sensor Data Collection*

The air quality sensor data is collected and stored firstly in a physical Gateway - Meshlium. The data transmission between the air quality sensors and the Gateway is secured through either HTTPS protocol (for the OAQ station which uses 4G communication) or through several encryption libraries which ensure that no third-party devices can connect to the sensors network (for the IAQ station which uses WiFi communication).

Afterward, is performed the data acquisition from the Gateway to the storage component with the help of an MQTT adaptor which allows a third party software component to subscribe to a specific data flow.

The MQTT data packets can be secured by implementing either Client ids or Usernames and passwords, thus the MQTT broker can require a valid username and password from a client before a connection is permitted. The username/password combination is transmitted in clear text and is not secure without some form of transport encryption. However, it does provide an easy way of restricting access to a broker and is probably the most common form of identification used. Plus, the username used for authentication can also restrict access to certain topics.

#### *Air Quality Sensor Data Storage and Access*

IAQ Sensor Data is stored in Azure Table Storage, in a storage account that also stores raw + processed motion camera data. Access to the data storage can be done in two ways; a security key-based access to the whole storage account, or a Shared Access Signature (SAS) based access to specific tables and key ranges. SAS access may also be limited to a certain time period within the signature itself. Access to the keys (and, consequently, to the ability to create SAS tokens) is restricted to limited people.

PostgreSQL storage

### 5.1.8 Patients' feedback

#### *Self-Report Data Sensitivity*

As the data directly concerns the subjects' health, it is treated as sensitive, personal data.

The true identity behind the hashed person or room codes are not stored in the system, but are rather only known to the researchers.

#### *Self-Report Data Collection*

The data collection is done via a dedicated application on the teacher's work phone. The application uses GRPC over HTTP/2. The commissioning of the questionnaires is done via a subject specific PIN key when taking the application into use. An Invalid PIN code results in a failed commissioning, preventing any further data from being entered. The GRPC interface does not support querying of already collected data. The data is not encrypted over transmission.



### *Self-Report Data Storage and Access*

Self-report data is stored in a MongoDB database within a dedicated virtual machine in Azure. Direct access to the database is prevented via firewall rules. The data is fetched via a REST API, which is restricted to select IP addresses only via firewall rules.

Due to the sensitivity of the data, access to it is restricted to VTT researchers only.

## 5.1.9 Physiological Data

Physiological data is collected via a combination of a Fitbit Smart Watch and an Android phone that has a role of acting as a gateway for the watch. The collected data includes heart rate data, e.g., raw PPG (pulse plethysmograph), burned calories, physical activities (customized activities or general type – walking, running, etc.) and sleeping patterns.

### *Physiological Data Sensitivity*

Obviously, physiological data are of the utmost sensitivity. Thus, access to it is granted only to VTT researchers. The data is pseudo-anonymised via a person code. The mapping between person codes and actual persons is known only to researchers, and not stored in the system.

### *Physiological Data Collection*

The air interface between watch and phone is protected via built-in BT encryption. Transfer of data from the phone to Azure is done using (plaintext) GRPC over HTTP/2. Each user is authenticated via a JWT (JSON Web Tokens) token on login for each session. Querying the collected data is not possible via the GRPC interface.

OAuth2

### *Physiological Data Storage and Access*

The data is received via a dedicated server running on a VM in Azure. The server stores the data into Azure Table Storage within the same data center. Querying the data is possible via either a REST API on the server that uses API key or JWT token based authentication (based on request), or directly from Table Storage using a storage account key. Access to the storage account has to be explicitly granted via Azure RBAC. Access to the API key also needs to be explicitly granted.

## 5.2 Indoor air quality improvement at school

The *Indoor air quality improvement at school* pilot is carried out in a Finnish school. The details of the school are not disclosed outside the necessary research staff in order to keep the possibility of derived identification of classrooms, teachers, etc. via other (public) information sources low.

Three sources of data are used in the pilot: IAQ sensors in classrooms, self-report questionnaires given by teachers, physiological data from watches and the accompanying phone used by the teachers, and 3D motion camera data. Each data source is discussed in the following, with focus on three points: the nature of the data (with respect to GDPR), protection of data during collection and transfer phase, and protection and access to stored data during analysis phase.

### 5.2.10 IAQ Sensors

IAQ sensor data consists of temperature, humidity, atmospheric pressure, pressure difference (in vs out), CO<sub>2</sub> level, Air quality index (0..500), noise level (dBA) , PIR sensor value, door sensor information

(open vs closed, state counter) and lighting level (lux). The data is collected from three different types of sensors.

#### *IAQ Sensor Data Sensitivity*

The sensor data does not contain location information or any other information that could be traced to the location (classroom) where the sensor is. The information for connecting the sensor data to a room is not stored in the system, but is rather only known to the select research persons (and the teachers). Because of this, the raw sensor data can be considered anonymous. However, room data can in cases be considered personal data, e.g., if the presence of any person in a room can be determined, it could in cases be linked to the presence of a specific person based on time of day / schedule, frequent users of the room, etc.

The collected sensor data may indicate presence or no presence in the rooms. Thus, room data combined from IAQ sensors is treated as personal data in the pilot, even if pseudo-anonymized with a hashed room code. External access to combined sensor data with any notion to the location of the room, will be evaluated on a case by case basis, potentially requiring a Data Processing Agreement (DPA).

#### *IAQ Sensor Data Collection*

All IAQ sensor data is collected using Bluetooth and LoRa built-in security for the air interface. Azure tokens are used for authentication, and all data transmission is encrypted with SSL. Standard Azure IoT hub security measures are in place, e.g., time to live based rejection of packets received outside the specified transmission window.

#### *IAQ Sensor Data Storage and Access*

IAQ Sensor Data is stored in Azure Table Storage, in a storage account that also stores raw + processed motion camera data. Access to the data storage can be done in two ways; a security key-based access to the whole storage account, or a Shared Access Signature (SAS) based access to specific tables and key ranges. SAS access may also be limited to a certain time period within the signature itself. Access to the keys (and, consequently, to the ability to create SAS tokens) is restricted to limited people.

### 5.2.11 Self-Report Questionnaires

Self-report questionnaires are filled in by the participating teachers at least twice a day. The questionnaires consider physical and mental symptoms and state (e.g., cough, shortness of breath, heavy-headedness, trouble concentrating, etc.) The teachers may also spontaneously report any symptoms or discoveries on their own wellbeing, or their observations on the students' behavior during the day. The answers are pseudo-anonymized by a hashed person code and a hashed room code.

#### *Self-Report Data Sensitivity*

As the data directly concerns the subjects' health, it is treated as sensitive, personal data. The true identity behind the hashed person or room codes are not stored in the system, but are rather only known to the researchers.

#### *Self-Report Data Collection*

The data collection is done via a dedicated application on the teacher's work phone. The application uses GRPC over HTTP/2. The commissioning of the questionnaires is done via a subject specific PIN key when taking the application into use. An Invalid PIN code results in a failed commissioning,

preventing any further data from being entered. The GRPC interface does not support querying of already collected data. The data is not encrypted over transmission.

#### *Self-Report Data Storage and Access*

Self-report data is stored in a MongoDB database within a dedicated virtual machine in Azure. Direct access to the database is prevented via firewall rules. The data is fetched via a REST API, which is restricted to select IP addresses only via firewall rules.

Due to the sensitivity of the data, access to it is restricted to VTT researchers only.

### 5.2.12 Physiological Data

Physiological data is collected via a combination of a Polar M600 watch and an Android phone that has a dual role of both acting as a gateway for the watch, and also collecting certain statistics of application usage that can be associated with stress. The collected data includes heart rate data, e.g., raw PPG (pulse plethysmograph) and heart rate variability data, activity data via accelerometer, phone usage, and location data.

#### *Physiological Data Sensitivity*

Obviously, physiological data are of the utmost sensitivity. Thus, access to it is granted only to VTT researchers. The data is pseudo-anonymised via a person code. The mapping between person codes and actual persons is known only to researchers, and not stored in the system.

#### *Physiological Data Collection*

The air interface between watch and phone is protected via built-in BT encryption. Transfer of data from the phone to Azure is done using (plaintext) GRPC over HTTP/2. Each user is authenticated via a JWT (JSON Web Tokens) token on login for each session. Querying the collected data is not possible via the GRPC interface.

#### *Physiological Data Storage and Access*

The data is received via a dedicated server running on a VM in Azure. The server stores the data into Azure Table Storage within the same data center. Querying the data is possible via either a REST API on the server that uses API key or JWT token based authentication (based on request), or directly from Table Storage using a storage account key. Access to the storage account has to be explicitly granted via Azure RBAC. Access to the API key also needs to be explicitly granted.

### 5.2.13 3D Motion Camera Data

Motion camera data is used to track people's movements within the classroom. The purpose is to detect presence / absence of people in the classroom, to get an overall sense of restlessness / calmness of people in the classroom, and to an extent, detect stress-related behavior from people's movement. The camera and the accompanying software (within the classroom) detects human head sized and shaped objects within given height boundaries. The data is processed locally into individual x/y points for each detected object at approximately 10Hz. The x/y point data is later combined into motion tracks for detected objects. This processing is done outside of the school premises.

#### *3D Motion Camera Data Sensitivity*

The data does not identify objects (persons) or even the height of the detected object, as long as it fits within the height boundaries. There is no other data attached to the x/y point detections. The data may indicate the presence / absence of a person in the room at a given time. Thus, if the actual classroom is

known, the data may become personal data. The link between the classroom and the classroom code is not stored within the system and is only known to the researchers.

### *3D Motion Camera Data Collection*

The camera data is locally processed on a Raspberry PI that transfers the raw point data to Azure Blob Storage once a day. The connection is secured via SSL, with token-based authentication using a storage account specific key.

### *3D Motion Camera Data Storage and Access*

Both the raw point data and the track data derived from it are stored as files in Azure Blob Storage, which is protected by a storage account specific key. The storage account is the same as for the IAQ sensor data, so the same key grants access to both IAQ sensor data and the point/track data. Analysis of the data (from points to tracks) is done on a computer at VTT. The computer accesses Blob Storage using SSL-secured transfer and token-based authentication with the storage account key. Access to the data has to be explicitly granted via the storage account key.

## **5.3 Tracking of athletes with wearable sensors**

Because of its relation with sports, entertainment, health and economy, sport and even every sports category has been seen as an industry. From the sport industry point of view, tracking and managing the parameters affecting the sport such as physical condition, health, motivation, eating, drinking and sleeping for amateur and professional athletes operating in individual or team sports in Europe and in our country became important. The development of wearable devices has given rise to new fields such as data analytics, reporting, developing a recommender system, while removing the problem of data collection. Wearable sensors, sensors that enable measurement of environmental factors and data from mobile applications can be analyzed to give suggestions as health and life assistant to the athletes. Turkish consortium will work on development of ESTABLISH platform on the topic of athlete management and athlete assistant. In this sense, our project, which aims to close an important gap, will present innovations in the sense of solving the problems experienced with its goals. The main objective of Turkish use case of the ESTABLISH project; defining an IoT infrastructure that integrates different sensor technologies. This infrastructure will includes the several capabilities such that:

- Environmental factors, body basic state parameters will be collected and analyzed with machine learning and text mining methods.
- According to the results, the services and the applications that are able to provide guidance and direction can be developed.

The main components of the project are:

1. Extracting meaningful information by analyzing collected data from mobile applications and IOT devices.
2. Integration of semantic data using Big Data platform,
3. Content-aware adaptation and automation of the IOT infrastructure,
4. Development of a suggestion system based on the results of data analysis.
5. Development of mobile and web applications software for tracking data, accessing analysis results and tracking recommendations.

The following figure 3 shows the high level architecture of Turkish use case. The platform security will be provided with IAM (Identity & Access Management).

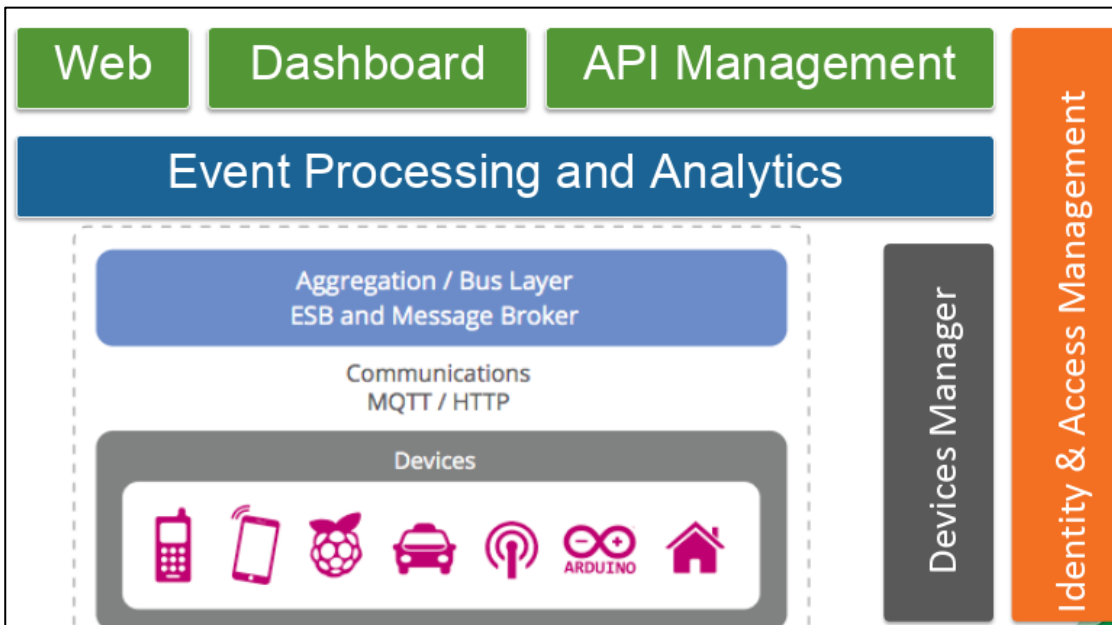


Figure 3. Establish Platform High Level Architecture

#### IAM for Establish Platform of Turkish use case

Establish Identity and Access Management (IAM) is a web service that helps you securely control access to the platform resources. IAM enables you to provide a seamless experience to all users of the platform. You use IAM to control who is authenticated (signed in) and authorized (has permissions) to use resources.

When you first create an Establish account, you begin with a single sign-in identity that has complete access to all Establish services and resources in the account. This identity is called Establish account *root user* and is accessed by signing in with the email address and password that you used to create the account. It is strongly recommended that do not use the root user for your everyday tasks, even the administrative ones. Instead, adhere to the best practice of using the root user only to create your first IAM user. Then securely lock away the root user credentials and use them to perform only a few account and service management tasks. An IAM solution can manage the complexities of connecting with most popular identity providers such as Salesforce, Twitter, Google. It does this while enforcing strict security policies with multi-factor and strong authentication.

## 6 Conclusion

The ESTABLISH project collects a lot of data that can be classified as personal data or even sensitive data. The variances between pilots are however evident and important to acknowledge since some data is purely open and environmental and do not require as complex management and storage processes as the pilots that collect data from humans. As ESTABLISH project is about smart health some health and physiological data gathering is vital for the success of the pilots and the project in general. It is not however illegal to collect such data, but to exploit it. This document has been an overview on what kind of data is collected and how this data is managed as well as the security measures that are applied to

the data. It is also important to note that possible risk factors in data becoming personal data and research subjects identifiable by the data have also been discussed. The ESTABLISH partners are committed to identify and prevent any risks concerning data collection and management and aim to produce the high security solutions and products that take into account data security as a built-in feature rather than an add-on.

This main objective of all pilots data security and protection is the implementation of reliable, secure data transport, storage and access. Special focuses is to ensure secure data management of security and privacy related issues where health, environmental or open data is concerned. The processes vary across pilots due to the data types they collect and the ways they have to protect it. However all processes can be narrowed down to four main elements: Access control, security of platforms, privacy by design mentality and incident management. To add a optional fifth one, all pilots have minimized the usage of data to the minimum level needed in their respective pilots.