D2.1 State-of-the-art, detailed use case definitions and user requirements

Deliverable ID:	D 2.1
Deliverable Title:	State-of-the-art, detailed use case definitions and user requirements
Revision #:	1.0
Dissemination Level:	Public
Responsible beneficiary:	VTT
Contributing beneficiaries:	All
Contractual date of delivery:	31.5.2017
Actual submission date:	15.6.2017

Table of Content

1. Introduction

This document (deliverable D2.1 State-of-the-art, detailed use case definitions and user requirements), presents the work done in work package (WP) 2 in the ESTABLISH project, during the first six months. The major goal of this WP2 is to provide definitions, requirements and specifications needed for the implementation of the ESTABLISH system. WP2 will collect and analyze the use case specific requirements to understand the motivation, interests, priorities and abilities of various stakeholders, establishing a context and concrete reference framework for the design and development of the technical solutions of the project. User needs and expectations towards defined use cases are explored via usability studies. Besides the requirements analysis, WP2 will codify the requirements in high-level technical specifications for the WPs 3 - 6. In those WPs, the requirements and specifications will be specified in more detail. Technical specifications will be documented in other deliverables.

The work was started together with the partners in the kick-off meeting held in March 2017. There we brainstormed the use cases in groups. The ideas were discussed and prioritized together and the owners were chosen for the use cases. The use cases were developed further in smaller groups with the companies participating the certain use case.

This deliverable is divided in six sections. After Introduction, the air quality related state-of-the-art is described in section 2 and technological state-of-the-art in Section 3. Section 4 presents the chosen use cases. Some of the use cases will be updated to this document later on, when the national funding decisions have been received. Use case requirements are presented in Section 5. Section 6 includes the conclusions and the next steps for WP2. All the references are listed at the end of the document. Use case specific technical specifications and general use case requirements are presented in Appendixes. Technical requirements of the use cases will be described more in detail in D2.2 Technical requirements and high-level specifications at the end of the year.

2. Air quality state-of-the-art

2.1 Indoor air quality effect on health and productivity

Indoor air quality effect on health

Since individuals spend majority of their time indoors, indoor air quality has clear effect on health, especially amongst more vulnerable people such as children and older people (Jones 1999). Indoor air pollutants (IAP) have the potential to cause transient morbidity, disability, disease, and even death in extreme cases. The likelihood that certain pollutant has an effect on individual's health depends on the individual's sensitivity to the contaminant, the contaminant concentration, the current state of their psychological and physical health, and the duration and frequency of exposure. There is growing body of literature demonstrating adverse health effects and reduced performance caused by poor indoor air quality (Tham 2016).

The sick building syndrome (SBS) is a medical condition where people in a building suffer from symptoms of illness or feel unwell for no apparent reason. The symptoms often worsen with the time spent in the building, and lessen or disappear after leaving the building. The common complaints of symptoms are headache, eye, nose, and throat irritation, fatigue, and dizziness and nausea. Also chest congestion (wheezing, shortness of breath, chest tightness), chills and fever, muscle pain, neurological symptoms (difficulty remembering or concentrating, feeling depressed, tension, or nervousness), inflammation, skin itching, and dry skin are among reported symptoms (Wikipedia 2017; Jones 1999).

Ventilation rate

The literature shows that higher ventilation rates in offices, up to about 25 l/s per person, are associated with reduced prevalence of sick building syndrome (SBS) symptoms and sick leave (Sundell et al. 2011). Sundell et al. (2011) conclude in their review that, although the available data was too limited for firm conclusions, low ventilation rates in schools were associated with increased absenteeism and more respiratory symptoms in schoolchildren. Maddalena et al. (2015) concluded that even though low ventilation rates had no statistically significant effect on perceived air quality (PAQ) or SBS, they had significant and independent negative impact on cognitive performance of healthy young adults.

Assessment of SBS symptoms and PAQ

In (Wargocki et al. 1999), subjective assessment of SBS symptoms was inquired with a questionnaire using visual-analogue scale (VAS) before and after exposure to air pollutants in the office environment. Table 1. represents twenty polar items used to assess SBS symptoms. Table 2. represents the questionnaire used to assess perceived air quality (PAQ).

Table 1. Visual-analogue scale items to measure SBS symptoms (Wargocki et al. 1999).

Right now I feel as follows:
Nose blocked - Nose clear
Nose dry - Nose running
Throat dry - Throat not dry
Mouth dry - Mouth not dry
Lips dry - Lips not dry
Skin dry - Skin not dry
Hair dry, brittle - Hair not dry
Nails brittle - Nails supple
Eyes dry - Eyes not dry
Eyes smarting - Eyes not smarting
Eyes aching - Eyes not aching
Eyes feel gritty - Eyes not feel gritty
Severe headache - No headache
Difficult to think - Head clear
Dizzy - Not dizzy
Feeling bad - Feeling good
Tired - Rested
Difficult to concentrate - Easy to concentrate
Depressed - Positive
Alert - Sleepy

Table 2. Questionnaire used to assess perceived air quality (Wargocki et al. 1999).

<table>
<tr><td colspan="2">Imagine that during your daily work you are exposed to this air.</td></tr>
<tr><td>How do you assess the air quality?

Pay attention to the dichotomy between acceptable and not acceptable</td><td>Clearly acceptable

Just acceptable
Just not acceptable

Clearly not acceptable</td></tr>
<tr><td>Assess odour intensity

No odour
Slight odour
Moderate odour
Strong odour
Very strong odour
Overpowering odour</td><td>Assess irritation in
Eyes Nose Throat
No irritation
Slight irritation
Moderate irritation
Strong irritation
Very strong irritation
Overpowering irritation</td></tr>
</table>

As labor costs in buildings exceed energy costs by two orders of magnitude (Wargocki and Wyon 2017), many works investigated how indoor air quality influences productivity of adults and schoolchildren. Two recent surveys in this area (Wargocki and Wyon 2017; Al Horr et al. 2016) summarized results of various works that reported notable influence of indoor air quality on performance of the test subjects, such as text-typing, proofreading, reading, memorization and mathematical tasks. Reduced productivity, in its turn, may have negative influence on one's health because a sense of self-efficacy has a significant impact on the level of stress (Bodys-Cupak et al. 2016), and stress is the second most frequent work-related health problem in Europe and a reason for at least half of lost working days in European enterprises (Alberdi, Aztiria, and Basarab 2016).

Ventilation rate

Wargocki and Wyon (Wargocki and Wyon 2017) summarized results of several lab and field studies and reported that in controlled conditions performance of text-typing task improved by about 1% for every two-fold increase in the outdoor air supply rate. In an open-ended test of creative thinking, subjects provided 10% more answers and more original answers at 10 L/s per person than at 3 L/s per person. In these experiments, the main pollutants were volatile organic compounds emitted from the new finishing materials that were present in the chamber where the experiments were carried out. In field studies the performance of call-center operators in Denmark improved by 6% when the outdoor air supply rate was increased from 2.5 to 25 L/s per person, but only when new bag filters had been installed in the ventilation system; with used bag filters an increase in the ventilation rate reduced measured performance by 8%. Similar results were observed in schools, for example, when the outdoor air supply rate increased, the speed at which the children performed four numerical and two language-based tasks improved significantly, and in the case of one numerical task the percentage of errors was significantly reduced. It could be shown that doubling the outdoor air supply rate would improve the performance of schoolwork in terms of speed by about 8% overall, and by 14% for the tasks that were affected significantly, with only a negligible effect on errors (Wargocki and Wyon 2017).

Thermal comfort

Thermal comfort also plays a significant role in occupant productivity. Dissatisfaction with thermal comfort leads to productivity loss (Al Horr et al. 2016). Studies indicate that temperature change from 18°C to 30°C can influence the performance of office occupants in tasks like typewriting, learning performance and reading. The temperature range from 21°C to 25°C is generally a stable range, but there is a decrease in occupant performance by 2% per 1°C increase in temperature in the range from 25°C to 30°C. Local thermal conditioning also influence occupants' comfort levels, and there is an evidence that personal control over thermal conditioning systems helps to improve productivity (Al Horr et al. 2016). However, research on the relationship between office tasks, temperature and productivity indicates that different kinds of mental work may have different optimal temperatures for productivity increase, e.g., one task may require colder temperature than another one (Al Horr et al. 2016).

Productivity evaluation

In lab studies productivity can be measured objectively, by calculating speed of performing various tasks and numbers of errors. Field studies usually employ subjective measures of productivity, such as self-reports of employees, despite that subjective feelings of the test subjects did not well correspond to objective productivity measures in the studies comparing these two approaches. The discrepancy between the two measures was not consistent in these studies, however. In one study with the insurance company self-estimates of performance showed no change in the period immediately following the move of the company to new premises, while objective measures showed a 30% decrease

in the rate at which insurance claims were being resolved in this period (Wargocki and Wyon 2017). In two studies in which both thermal and air quality conditions deteriorated or improved at the same time, self-estimated performance underestimated the magnitude of the effect on objectively measured performance in one experiment, but overestimated it in another experiment (Wargocki and Wyon 2017). In an experiment in which 99 subjects worked for 2 h under very different combinations of indoor environmental conditions, some good and some bad, self-estimated performance improved significantly by as much as 25% as conditions improved, while objectively measured performance improved non-significantly and by only 7% (Wargocki and Wyon 2017).

A decrease in self-estimated performance may therefore indicate no more than increased subjective dissatisfaction with working conditions; however, dissatisfaction reduces motivation and thus may have negative long-term effects not observed in the relatively short-term studies, described above. In addition, objective measures of productivity in real life offices are yet to be found. Currently productivity can be evaluated only indirectly, by counting the number of hours worked each week, safety rule violations, the number of grievances filed, and employee turnover (Al Horr et al. 2016). However, although e.g. absences from work are related to productivity, presence at work does not necessarily imply high productivity. For example, a person can be stressed, tired or unable to focus on her work due to interruptions. Hence productivity can be also indirectly evaluated by detecting personal stresses and estimating either personal concentration on the work tasks or its reverse - so-called interruptibility (a mental state in which interruptions are the least harmful) because there is an evidence that moments of low mental load and switching between tasks are the most suitable for interrupting one's work (Züger and Fritz 2015). Both stresses and interruptibility can be detected via physiological sensors, listed in section 2.3 3.2 Objective health monitoring with sensors, and via behavior monitoring, such as evaluation of postures, usage of computer keyboard and mouse, mobile phone data usage, facial expressions, audio analysis etc. (Züger and Fritz 2015; Alberdi, Aztiria, and Basarab 2016).

2.2 Outdoor air quality

The claim of environment influence in well-being and human health is dated back two millennia when Hippocrates is his text *On Airs, Waters and Places* introduces several ideas that do retain currency. For example, the simple message that good health is unlikely to be achieved and maintained in poor environmental conditions is enduring (*apud*. George Morris and Patrick Saunders 2017). The environmental compartments identified by Hippocrates i.e. water, soil, air, and food still have operational and conceptual relevance.

Good quality natural environments can bring many benefits to physical, mental and social well-being. However, environmental degradation - such as that caused by air, soil and water pollution with their specific components like noise, radiation, chemicals or biological agents - have an overall negative impact on the well-being of individuals and on long term affects the health condition. Despite substantial improvements in recent decades, environmental health / justice remains a challenge for governments around the world.

Due to social, economical, environmental, lifestyle trends and fast changing technologies and unequal distribution of environmental and socio-economic conditions specific diseases and affections emerge. In the figure below you may see the effervescence of pollution and some of the consequences of environmental degradation.

Air pollution is among the most investigated environmental threats to health and well-being. Royal College of Physicians publishes a report in 2016 on the lifelong impact of air pollution assessing that Over the next half century a different, more insidious form of air pollution appeared, linked to the emissions from the ever-increasing number of motor vehicles and other forms of transport on our roads,

rail and seas that are dependent upon combustion of petrol and diesel fuels. Added to this chemical onslaught is the effect of indoor air pollution from workplaces and the fittings, furnishings, heating and cooking in our homes. (RCP, Every breath we take – the lifelong impact of air pollution, 2016)

Yet the industry continues to contribute significantly to emissions of air pollutants and greenhouse gases in Europe. In 2012, the industry was responsible for 85% of the emissions of sulfur dioxide (SO2), 40% of the emissions of nitrogen oxides (NOX), 20% of the emissions of particulate matter (PM2.5) and non-methane organic volatile compounds and 50% of the emissions of greenhouse gases in AEM-33 countries (EEA). Research points to the following effects of the changes in gas concentrations on the human body:

Air pollution consists in the change of the chemical composition of air in terms of the ratio of its constituent substances and the emerging ones, with harmful effects for the environment as a whole or some of its components.

Sources of air pollution include: natural sources (barren, dry soil which produces dust; plants and animals, from which pollen, lint, hair, feathers, fungal spores, etc. come, volcanic eruptions, which is a considerable source of volcanic ash, gases (CO_2, SO_2, H_2S, etc.) and water vapor; space, from which cosmic dust comes and artificial sources (sources related to the burning of fuels in household activities, in the energy industry; sources based on industrial processes such as those in the chemical, steel, metallurgic industries, materials of construction; mobile sources, including all means of transport spreading pollutants over long distances. Thus, in Central and Eastern Europe mobile sources (especially road traffic) are responsible for about 30-60% of the emissions of carbon monoxide, between 35 and 95% of the emissions of lead, and less than 5% of the emissions of sulfur dioxide. The vehicles also emit small amounts of toxins and carcinogens such as benzene and aldehydes.

According to the report (2015) entitled “Europe's Environment - state and prospects", prepared by the European Environment Agency (EEA), in collaboration with the European Network of Environment Information and Observation (Eionet) and the European Commission, on the current situation and environmental prospects, Romania aims to improve air quality and the sustainable management of water in compliance with EU rules (http://www.eea.europa.eu/soer-2015/synthesis/mediul-european-2013-starea-si#tab-data-visualisations).

The report stresses the fact that industrial emissions of pollutants were reduced but still causing considerable damage every year. Just like the energy and transport sectors, the European industry brings a number of benefits, but also costs to society. Besides the production of goods and services, this sector generates a substantial number of jobs, profits and tax revenues. At the same time, however, the industry contributes significantly to emissions of several important air pollutants and many greenhouse gases, causing widespread damage to the environment and people's health.

Yet the industry continues to contribute significantly to emissions of air pollutants and greenhouse gases in Europe. In 2012, the industry was responsible for 85% of the emissions of sulfur dioxide (SO2), 40% of the emissions of nitrogen oxides (NOX), 20% of the emissions of particulate matter (PM2.5) and non-methane organic volatile compounds and 50% of the emissions of greenhouse gases in AEM-33 countries, like Figure 1. shows.

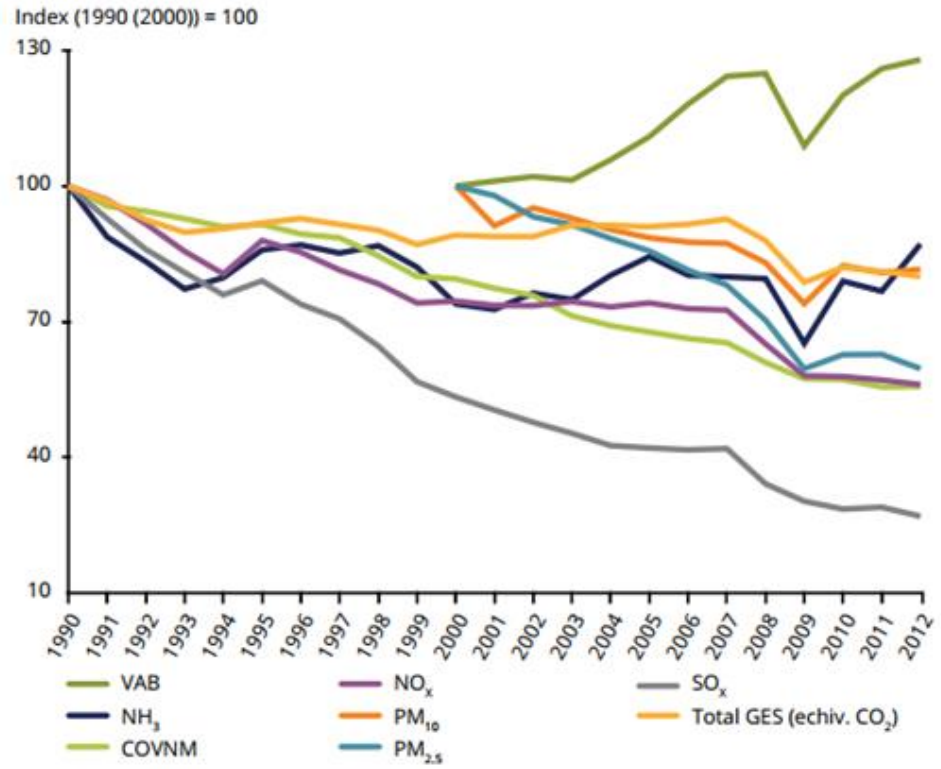

Figure 1. Emissions of air pollutants and greenhouse gases in Europe during 1990-2012 (EEA).

According to EPA (Environmental Protection Agency) AQI- Air Quality Index is an index for reporting daily air quality. EPA calculates the AQI for five major air pollutants regulated by the Clean Air Act: ground-level ozone, particle pollution (also known as particulate matter), carbon monoxide, sulfur dioxide, and nitrogen dioxide, see Figure 2. For each of these pollutants, EPA has established national air quality standards to protect public health.Ground-level ozone and airborne particles are the two pollutants that pose the greatest threat to human health in this country (https://airnow.gov/index.cfm?action=aqibasics.aqi).

Air Quality Index (AQI) Values	Levels of Health Concern	Colors
When the AQI is in this range:	*..air quality conditions are:*	*...as symbolized by this color:*
0 to 50	Good	Green
51 to 100	Moderate	Yellow
101 to 150	Unhealthy for Sensitive Groups	Orange
151 to 200	Unhealthy	Red
201 to 300	Very Unhealthy	Purple
301 to 500	Hazardous	Maroon

Figure 2. AQI categories.

Research points to the following effects of the changes in gas concentrations on the human body:

- **Nitrogen dioxide** is known to be a very toxic gas for both humans and animals (the degree of toxicity of nitrogen dioxide is 4 times higher than that of nitric oxide). Exposure to high concentrations can be fatal, and at low concentrations it affects the lung tissue. The population exposed to such pollutants can have difficulty breathing, respiratory irritation, and lung dysfunction. Long-term exposure to even low concentrations can damage lung tissues resulting in emphysema. Children are the most affected by exposure to this pollutant.
- **Carbon monoxide** is a toxic gas, fatal in high concentrations (concentrations of around 100 mg / m3) by reducing the oxygen carrying capacity of the blood, with consequences on the respiratory system and cardiovascular system. At relatively low concentrations: it affects the central nervous system; it modifies the heart rate, thus decreasing the volume of blood distributed in the body; it reduces visual acuity and physical ability; exposure over a short period can cause acute fatigue; it can cause difficulty breathing and chest pain for people with cardiovascular disease; it causes irritability, headaches, rapid breathing, lack of coordination, nausea, dizziness, confusion, reduced ability to concentrate. The population most affected by carbon monoxide exposure is represented by: children, the elderly, people with respiratory and cardiovascular diseases, anemic people, and smokers.
- **Particulate matter** is a complex mixture of very small particles and liquid droplets. An important problem is represented by the particles with an aerodynamic diameter less than 10 micrometers, passing through the nose and throat and penetrating into the pulmonary alveoli causing inflammation and intoxication. They affect mainly people with cardiovascular and respiratory diseases, children, the elderly and the asthmatics. Children younger than 15 years inhale more air, and therefore more pollutants. They breathe faster than adults and tend to breathe more through the mouth, practically bypassing the natural filters in the nose. They are especially vulnerable because their lungs are not developed and lung tissue which develops in childhood is more sensitive. Dust pollution worsens asthma symptoms such as cough, chest pain and difficulty breathing.

Exposure to increased levels of air pollution and deteriorated indoor climatic conditions has a severe impact on health and well-being. It leads to increased medication intake and a higher number of hospital admissions, especially for vulnerable groups such as elderly, patients with respiratory or cardiovascular problems, and to a lesser extent also employees in office environments and children.

Beyond the common recommendations on improving our lifestyle via exercising, eating healthier and relaxing more, access to all this resources is not the same for everyone considering social, geographical membership (inequities).

The number of those doing outdoor physical activities (jogging) or in specially designed rooms is growing. Butt exercising outdoor means also exposure to allergens. Together with those who intend to maintain their fitness and prevent obesity, there are people who have prescriptions for physical activity for the recovery/restoring of motor functionalities, both outdoors and indoors.

Assisting persons that exercise outside via cutting-edge devices and providing access date in other to help them monitor their overall state of health and impact of external factors is one o the major challenges raised by modern society.

Sustainable transportation

Over the last ten years, many cities have been experiencing a "back to the city" movement. Younger and older generations alike are recognizing value in the diversity of social and cultural experiences, economic opportunities and increased quality of life that cities offer. As a result, more and more people

across demographic lines are choosing to relocate to urban areas. To accommodate this movement, some governments increased service provision - be it infrastructure services, housing, public safety or recreational amenities - in order to meet increasing demands from a diverse group of constituents. As populations shift towards urban centers, many governments are re-examining current transportation systems to determine how to accommodate a growing and socioeconomically diverse population, while seeking to minimize environmental pollution and urban congestion. Within this context, sustainable transportation approaches, such as public transit and comprehensive bicycle and pedestrian networks, become a critical part of a city's growth strategy and its ability to effectively meet the needs of its residents.

Sustainable transportation options can serve to attract and accommodate the varied needs and desires of urban residents and are often part of a long-term strategy that includes the integration of various systems, both across geographies and modes of travel. A sustainable transportation system is one that allows the basic access and mobility needs of all individuals to be met safely; is affordable; provides multiple transport options and supports both public health and a vibrant economy; limits emissions; and minimizes consumption of non-renewable resources. Although personal vehicles can be part of a sustainable transportation system, there is an increased emphasis on public, non-motorized, and multi-modal transportation options. However, we did not find in the literature transportation systems that respect also the citizen's sensitivity levels and their requirements / constraints (Energy-efficiency, CO_2 and NO_2 sensitivity, queue lengths at public services availability etc.) that we would like to handle in this project.

3. Technological state-of-the-art

3.1 Self-awareness and self-adaptivity to cope with uncertainty

An inevitable factor in the systems that monitor environment, process the data and act upon them is the uncertainty. This comprises internal uncertainty that pertains to the system's operation and health and the external uncertainty, which stems from physical processes that are not anticipated by the systems and the human-in-the-loop. In order to provide reliable service, this uncertainty has to be coped with and the system has to be built to proactively mitigate its effects. The typical solution for dealing with uncertainty is to build on autonomic computing, most notably on the concept of the MAPE-K loop. This represents periodic process of Monitoring, Analysis, Planning and Execution, which all operate around Knowledge. The purpose of the MAPE-K loop is to observe the operation of the system and reconfigure/adapt the system to comply with a situation in hand (typically changing and not completely anticipated at design time due to uncertainty). The concept of the MAPE-K loop is then extended by several reference architectures for adaptive systems (Huebscher and McCann 2008). This is also embodied by frameworks such as Rainbow with adaptation languages such as Stich (Cheng, Garlan, and Schmerl 2012) and formalized by formalisms such as Forms and ActivForms. The self-adaptivity is typically combined with self-awareness, which pertains both to the (i) introspection of systems operation and (ii) runtime awareness of system's high-level goals, their satisfaction (also referred to as "requirements reflection") and strategies to solve them.

The (i) introspection of system's operation includes scalable methods for monitoring of systems functional and non-functional properties (quality of services – QoS). This comprises a plethora of models and methods for runtime monitoring, typically bound to a specific metric – e.g. latency, bandwidth utilization, energy consumption, etc. A challenge here is the ability of observing the system while not disturbing it or negatively influencing its operation (e.g. by minimizing the CPU and bandwidth needed for collecting the performance data).

The (ii) runtime awareness of system goals, this is tackled by methods of goal-based requirements engineering (e.g. Tropos and KAOS (Lamsweerde 2008) and extended to runtime self-adaptivity based on high-level goal by approaches such as FLAGS and IRM-SA (Gerostathopoulos et al. 2015), which allow multilevel monitoring of system's goals and decisions about their satisfaction.

3.2 Objective health monitoring with sensors

Personal activity and physiological monitoring technologies have become common in every-day life over the last decade. ***Activity monitoring devices*** are used to measure, e.g., activity type, level of activity, energy consumption, physical functioning, gait and fitness. There are a variety of sensors used for movement monitoring, e.g., accelerometers, gyroscopes, goniometers, force sensors, pressure sensors, altimeters, and location GPS (Rodgers, Pai, and Conroy 2015; Piwek et al. 2016). Often activity monitoring device uses combination of several sensors, e.g. Inertial Measurement Unit (IMU) that contains accelerometer, gyroscope, and often also magnetometer. Advances in microelectromechanical systems (MEMS) have resulted in sensors that are small, low-cost and operate in a wireless manner, and nowadays activity sensors can be found in most of the smart phones enabling development of variety of applications for movement and health monitoring. In their simplest form, the activity monitors differentiate between activity and rest, often assuming human motion entails activities with a repetition rate above 0.1 Hz. However, wearable sensors may also identify different static positions (sitting, lying, and standing), postural transitions and dynamic activities, such as walking, climbing stairs and cycling (Shany et al. 2012).

Physiological monitoring devices measure, e.g., heart rate (HR), heart rate variability (HRV), breathing rate, blood oxygen saturation, blood pressure, electroencephalography (EEG), electrocardiography (ECG), electromyography (EMG), skin temperature and skin conductivity (Rodgers, Pai, and Conroy 2015; Gjoreski et al. 2016). Electrocardiogram (ECG) signal provides information about heart functions. Heart rate variability (HRV) is the fluctuation in the time between the successive RR-intervals in ECG signal. HRV represents autonomic modulation of the heart and decrease in HRV has been associated with severe health outcomes, even increased mortality (Stein and Kleiger 1999). HRV expresses sympathetic and parasympathetic nervous system activity and it may be used to analyze individual physiological stress and recovery originated from lifestyle-related behaviors and environmental effects (Firstbeat Technologies Ltd. 2014). Also, electro-dermal activity (EDA) that indicates skin sweating activity has been used to detect stress of a person (Sandulescu et al. 2015). EDA can be measured, e.g., as skin conductivity or galvanic skin response (GSR). Furthermore, variation in skin temperature may indicate medical stress that precedes health outcomes, such as stroke, heart attack and shock (Mukhopadhyay 2015).

Physiological sensors require consistent skin contact to measure the biological, chemical or physical phenomena, which poses a challenge for the technology - how to maintain consistent contact over extended periods and under different conditions. From the healthcare point of view the challenge is - how to achieve high sensitivity and specificity for detecting abnormal events in real-time. Capacitive sensors are typically used to measure biopotentials for EEG, ECG and EMG (Rodgers, Pai, and Conroy 2015). Recently, new optical detection techniques have emerged as alternatives to traditional heart rate monitoring devices that require a separate chest strap. Wrist-worn monitors measure the heart rate via photoplethysmography (PPG) that optically detects changes in skin blood volume by using a light-emitting diode (LED). The advantage of wrist-worn devices is that they are unobtrusive and suitable for continuous and long-term wear. However, they might not reach the same level of accuracy when compared to electrode-based chest strap devices, especially during exercise (Wang et al. 2017). Furthermore, recent advance in material science and fabrication techniques of flexible electronics have brought up new types of health monitoring devices that can be adhered directly to the skin (e.g. patch sensors), and clothing integrated sensing devices (Rodgers, Pai, and Conroy 2015). Currently, the line between consumer devices and medical devices begins to blur and one single device can measure range of physiological variables and medical risk factors (Piwek et al. 2016). In addition to wearable monitors, there are devices such as Body Cardio scales from Withings (www.withings.com) that can assess cardiovascular health besides weight and body composition. Table 3. lists a few example solutions for health monitoring.

Table 3. Examples of companies offering health-monitoring devices.

Company	Products	Signals and variables
Mega Electronics Ltd. http://www.megaemg.com/products/	eMotion Biolink eMotion ECG eMotion Faros eMotion HRV Scanner	ECG heart rate variability (HRV) EMG EEG stress recovery general well-being
Withings https://www.withings.com/eu/en/	Steel HR Withings GO Withings Pulse O_x Body Cardio	heart rate activity sleep steps distance elevation calories blood oxygen level weight, body composition, pulse wave velocity
OURA https://ouraring.com	OURA ring	sleep activity skin temperature resting heart rate
Polar https://www.polar.com/uk-en/products	Several products with different features, heart rate monitors, fitness trackers, GPS sport watches e.g. Polar A360, Polar Loop, M600, V800	activity sleep steps heart rate (wrist-based + chest strap) calories speed distance GPS
Fitbit http://www.fitbit.com/fi	Fitness trackers, e.g. Fitbit One, Fitbit Flex 2 Fitness trackers with HR, e.g. Fitbit Alta HR Fitness watches, e.g. Fitbit Surge	activity steps calories sleep heart rate (wrist-based) GPS
Muse http://www.choosemuse.com/	muse, brain sensing headband	EEG
Alivecor https://www.alivecor.com/	Kardia	ECG (fingers)
Zensorium http://www.zensorium.com/	Being Tinke	Being: steps, calories, heart rate, distance, mood, sleep Tinke: heart rate, respiratory rate, blood oxygen level, HRV, stress
Garmin http://www.garmin.com	Several products for activity and sports monitoring, e.g. Vivo series, Fenix series	heart rate (wrist) GPS activity steps stress barometric altitude sleep
Samsung http://www.samsung.com/us/mobile/wearables/	Samsung Gear	Accelerometer, Barometer, Gyro Sensor, Heart Rate Monitor, Ambient Light Sensor
Apple https://www.apple.com/watch/	Apple Watch	heart rate (wrist) activity sleep GPS
Empatica https://www.empatica.com/e4-wristband	Empatica E4	PPG (wrist), accelerometer, GSR, skin temperature
VitalConnect https://vitalconnect.com/	VitalPatch	ECG, HR, HRV, BR, skin temperature, body posture, fall detection, activity

Indoor air quality effects on physiological measures

Especially indoor air temperature has been shown to have an effect on perceived air quality, SBS symptoms, physiological responses, and human performance (Lan et al. 2011; Liu, Zhong, and Wargocki 2017). Lin et al. (2011) showed that average heart rate, respiratory ventilation rate, and end-tidal partial CO_2 ($ETCO_2$) were higher when indoor air temperature was 30°C when compared to those measured at 22°C, suggesting increased metabolic rate in higher temperature. On the contrary, the SPO_2 measurements indicated that arterial oxygen saturation was significantly lower at 30°C compared with 22°C. Lower SPO_2 was associated with intensity of SBS symptoms, task performance, increased fatigue symptoms and decreased cognitive functions. Liu et al. (2017) studied effects of indoor air temperature increase from 26°C to 35 °C and CO_2 concentration on physiological responses, subjective ratings and cognitive performance of 12 healthy young adults. The measured physiological parameters were skin temperature (six locations on body), eardrum temperature, heart rate variability, ventilation rate, blood pressure, end-tidal partial CO_2, arterial blood oxygen saturation and body weight. According to their results, increasing CO_2 concentration did not cause changes in response variables. However, they conclude that the measured eardrum temperature, skin temperature, heart rate, respiratory ventilation rate and body weight loss increased at the higher temperature. Heart rate variability, measured as the percentage of adjacent inter-beat intervals differing by > 50 m, and arterial oxygen saturation were lower in higher temperature (Liu, Zhong, and Wargocki 2017).

3.3 Sensors available for indoor air quality monitoring

The main factors in the indoor environment affecting well-being and health may be classified as follows (European Commission 2007; Sisäilmayhdistys):

- Chemicals. Chemicals for intended use or unintentional emissions from different sources including e.g. gas emissions from different materials, volatile organic compounds and tobacco smoke.
- Particulate matter. Sources of particle pollution include, for example, dust, asbestos, microbes, and allergens such as mold. Typically, the smaller ones come from the outside air and are caused by traffic etc.
- Physical factors include such as humidity, temperature and ventilation. Physical factors are not pollution but affect comfort of indoor environment.

Recommended values for each factor are given in standards and other regulations in order to maintain health and comfort (Persily 2015; WHO 2010). There are hundreds of gases and chemicals that might be detected in indoor environment but typically only a few are monitored (WHO 2010):

- Volatile organic compounds (VOC): The volatile organic compounds are usually monitored with TVOC (Total Volatile Organic Compounds) sensors that give combined value of VOCs in air. VOC measurement sensors can be, for example, Photoionization Detectors (PID) or Gas Sensitive Semiconductor (GSS). PID type VOC sensors are more resilient and can detect a wider variety of VOCs.
- Formaldehyde is a colorless gas of which indoor sources may be combustion processes such as smoking and heating. Major sources in non-smoking environments appear to be new building materials and consumer products that can emit formaldehyde even several months, particularly in conditions with high relative humidity and high indoor temperatures.
- Radon gas is a major contributor to the ionizing radiation dose received by the general population. House with poorly sealed foundations, built on high-permeability ground and with several entry points for soil gas may draw more than 10% of their indoor air from the soil. Even if the soil air has only moderate levels of radon, levels inside such houses may be very high. Water supplies can also contribute to indoor radon levels.

- CO Carbon monoxide (CO) is a colorless, non-irritant, odorless and tasteless toxic gas. It is produced by the incomplete combustion of carbonaceous fuels such as wood, petrol, coal, natural gas and kerosene.

Particulate matter detectors are able to count and size particles in the air. Traditionally, particles have been divided into coarse particles (PM10) that are between 2.5 and 10 micrometers in diameter and fine particles (PM2.5) that are less than 2.5 micrometers in diameter. There are three types of real-time instruments that are normally used for counting particles: photometers, optical particle counters and condensation particle counters (TSI 2013). Usually they are intended for professional use (e.g. power plants) but there are also coming low cost particle counters on the market (Prabakar, Mohan and Ravisankar 2015) that are more consumer products.

Sensor measuring physical factors include (TSI 2013):

- Air velocity. Ventilation is one of the most important contributing factors to indoor air quality. Low ventilation rates have been associated with health and work performance. Air velocity is a good indicator that air is being appropriately distributed throughout the building.
- Carbon Dioxide (CO_2). Another good measure of proper ventilation is the level of CO_2. Carbon dioxide is a normal by-product of respiration, combustion and other processes.
- Temperature is one of the basic measures of air quality. Extreme indoor temperatures are a serious health hazard.
- Humidity can be measured in several ways, for example, relative humidity, humidity ratio and absolute humidity. In addition to affecting the thermal comfort of occupants, dampness or moisture may accumulate into the building structures of finishing materials that may lead to mould and microbial growth.

3.4 Networking technologies for building monitoring applications

The choice of network architecture is one of the most important decisions with IoT deployment. Network technologies can be classified as *wired* or *wireless* (see Table 1). The main advantage of a wireless network over a wired one is that users and devices can move around freely within the area of the network and get an internet connection, while wired connections are still useful for relatively more reliable, secured and high-volume network routes. (Kaur 2014). The selection of technology is a tradeoff between factors such us Quality of Service (QoS), battery life, latency, network coverage and range, deployment model, and cost (Sinha et. al 2017).

Figure 3. describes the IoT Connection technologies (Postcapes 2014). When data have to be transferred over short distances, devices can use wireless personal area network (PAN**)** technologies. A local area network (LAN) connects networked devices over a bigger area, such as office building, school or home. When data has to be transferred over large physical distance, wide area network (WAN) technologies are used. A WAN is geographically spread collection of LANs, which are connected through routers. The Internet is an example of the largest WAN spanning the Earth.

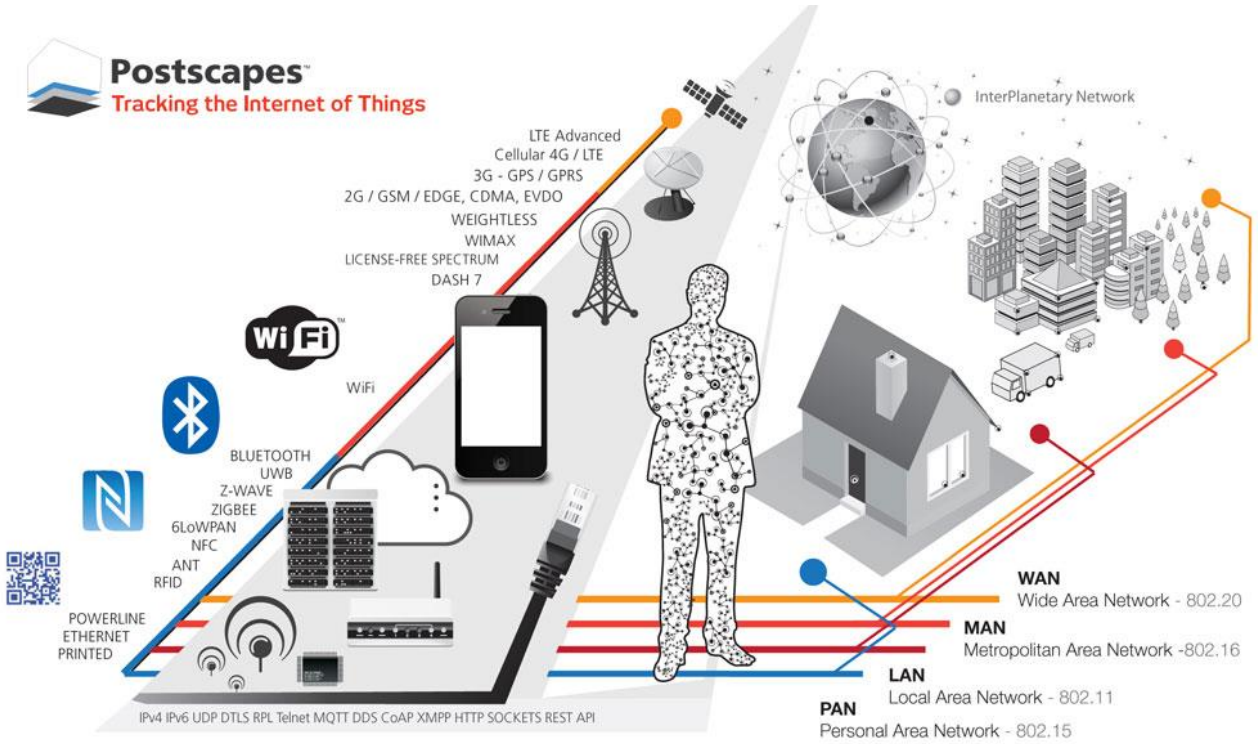

Figure 3. IoT Communication Technologies (Postcapes 2014)

Currently, the most essential short-range wireless network technologies can be narrowed down to the following:

- **Bluetooth, Bluetooth LE and Bluetooth Mesh** (Bluetooth): Bluetooth is based on IEEE 802.15.1 standard and targeted for exchange of data over short distance. The Bluetooth Low Energy (BLE) extension was introduced in core specification v4.0 and makes possible to combine low power cell coin devices, low duty cycle and low latency with Bluetooth to build a robust and secure system. BLE is well suited for energy-constrained devices. The main drawback for using Bluetooth as an IoT platform is the limited number of connections that the phone or tablet can support. The Bluetooth Mesh, an extension of BLE, is an answer to this problem. Mesh networks allow signals to be repeated from one node to another across a network. A central hub or gateway is not needed unlike in many other standards.
- **NFC** (NearFieldCommunication): Near Field Communication (NFC) is a set of short-range wireless technologies requiring a distance of 10 cm or less. NFC involves an initiator and a target; the initiator actively generates an RF field that can power a passive target. This enables NFC targets to take very simple form factors such as tags, stickers, key fobs, or cards that do not require batteries.
- **RFID** (Want 2006): Radio Frequency Identification (RFID) is a contactless identification system based on radio technology. Information is stored in electronic form in RFID tag attached to the objects to be identified and can be read with a RFID reader over a short distance. The readers generally transmit their observations to a computer system running RFID software or RFID middleware. Tags can be manufactured in various sizes and shapes, from inexpensive passive tags to miniature implant tags and active tags with sensor capabilities. Applications for RFID are diverse as a tag can be embedded into any object in order to track and manage inventory, assets, people, cars, computer equipment, books, mobile phones etc.
- **Wi-Fi** and **Flow Power Wi-Fi** (W-Fi Alliance): Wi-Fi refers to the IEEE 802.11 set of standards for wireless networks and has become the standard technology for wireless networking. Wi-Fi

provides rather high bandwidth capacity, but is typically more expensive and consumes more power than for example ZigBee. Higher-cost, lower-power Wi-Fi devices have “sleep mode” when not transmitting data. Low Power Wi-Fi with batteries can be used for remote sensing and control applications (Dobkin 2009).

- **ZigBee** (ZigBee): ZigBee is a protocol suite developed by the ZigBee Alliance for low-cost, low-power wireless communications, often used in wireless sensor networks and in building automation. ZigBee builds on top of the IEEE 802.15.4 standard that defines the two lowest layers of the stack, medium access control and physical layers. On top of these two layers, ZigBee standard adds a network layer and an application-supporting framework. In addition, ZigBee defines its own security architecture. Compared to Wi-Fi and Bluetooth, ZigBee uses very little power and is more cost efficient (Gomez 2010).
- **6LoWPAN** (Shelby 2011): IPv6 over Low power Wireless Personal Area Networks (6LoWPAN) is targeted to applications that need wireless internet connectivity at low data rates. Internet Engineering Task Force 6LoWPAN WG has defined an adaptation layer that enables IPv6 over IEEE 802.15.4 based networks. 6LoWPAN provides an important capability for future sensor networks because of its high compatibility with the present Internet infrastructure, allowing the connection of these networks with other IP networks. 6LoWPAN is typically used in sensoring and controlling indoor conditions, such as temperature, humidity and light.

The most commonly employed wide range wireless network technologies are the following:

- **Cellular technology** (3GPP)**:** Cellular technology, such as 3G, 4G (LTE, LTE-A), is what mobile phone networks and based on, *i.e.*, open technology used for transmitting mobile voice and data. Basically, cellular technology refers to having many small-interconnected transmitters as opposed to one big one. Data transfer rates and energy requirements are two key considerations when selecting a network technology for a given application. Technologies, such as 4G (LTE, LTE-A) and 5G are favorable for IoT applications, given their high data transfer rates. NB-IoT is a new IoT technology that is integrated into the LTE standard but it can be considered as a new air interface because it removes many features of LTE, including handover, measurements to monitor the channel quality, carrier aggregation, and dual connectivity. NB-IoT has been kept simple in order to reduce device cost and minimize battery consumption. It is especially suitable for applications that require high QoS and low latency (Sinha 2017).
- **LoRaWAN** (LoRa): A Low Power Wide Area Network (LoRaWAN) is a new network technology intended for wireless battery operated things in regional, national or global network. LoRaWAN targets key requirements of the Internet of Things such as secure bi-directional communication, mobility and localization services. LoRaWAN network architecture is typically laid out in a star-of-stars topology. LoRa is especially appropriate for the low cost applications (Sinha 2017).
- **SigFox** (SigFox)**:** SigFox provides global LPWA (Low Power Wide Area) network that offers Ultra Narrow Band radio modulation, lightweight protocol for small messages and small payload with star network architecture. SigFox offers a software based communications solution, where all the network and computing complexity is managed in the Cloud, rather than on the devices. This approach reduces energy consumption and costs of connected devices
- **Weightless** (Weightless): Weightless technology delivers wireless connectivity for low power, wide area networks specifically designed for the Internet of Things. Weightless can operate in both sub-1GHz license exempt and licensed spectrum. Weightless is an LPWAN technology Open Standard enabling long range, long battery life and ultra-low cost IoT products.
- **WiMax** and **WiMAX 2** (WiMax): Worldwide Interoperability for Microwave Access (WiMAX IEEE 802.16) is a standard based on wireless technology that provides high-throughput broadband connections over long distances. WiMAX can be used for a number of applications, including "last mile" broadband connections, Hotspots and for an alternative to 3G. WiMAX is emerging as

the future of broadband communication, providing affordable high-speed, long range, wireless performance, and as an alternative to wired landline networks. WiMAX technology also provides enterprise connectivity for business and remote users to connect to corporate networks and the Internet at far greater ranges than current wireless technology allows. WiMAX 2 is the latest technology in the WiMAX family.

The main challenges are related to adoption of network technologies are associated with security, network interconnections and power consumption. Security will be a major concern wherever networks are deployed. Data transmission can be relied on several networks, such as Ethernet, cellular or other wireless networks, and different network technologies are often connected with gateways (Latvakoski 2014; Iivari 2014). In addition to security, these interconnections pose also challenges by adding complexity (Ma 2011). In order to tackle the problem networked devices have with power consumption, energy-efficient networking has been and is still explored in research communities (Feng 2013).

3.5 Data Visualization Frameworks

This point describes the state of the art in the representation and display of critical information. Information and data represented in the best and fastest way to capture the message. The visualization of data is the process of searching, interpreting, contrasting and comparing data that allows an in-depth knowledge and detail of them in such a way that they become information comprehensible for the user.

The right way to display these views is one that makes up a collection of assembled resources to create a single unified visual display (dashboard). Dashboards often provide at-a-glance views of KPIs (key performance indicators) relevant to a particular objective.

Dashboards can be broken down according to role and are either strategic, analytical, operational, or informational. Strategic dashboards support managers at any level in an organization. Dashboards of this type focus on high-level measures of performance, and forecasts. Strategic dashboards benefit from static snapshots of data (daily, weekly, monthly, and quarterly) that are not constantly changing from one moment to the next. Dashboards for analytical purposes often include more context, comparisons, and history, along with subtler performance evaluators. Dashboards for monitoring operations are often designed differently from those that support strategic decision-making or data analysis and often require monitoring of activities and events that are constantly changing and might require attention and response at a moment's notice.

Data visualization selection

In the large majority of dashboard projects, we have found that key performance indicators (KPIs) fall into a handful of core categories, which have specific implications for visual analytics. As such, we have mapped visualization guidelines and tips to the different varieties of KPIs, along with examples.

While this section starts out with the basics, a further analysis should be done into each one, as each view is considered valid for the panel in question.

Quantities

One may think of these as counts or absolute measures, such as the number of units sold, time taken to complete an examination, or tons of freight carried by a train car.

There is a wide variety of options for visualizing basic quantities. Bar or column charts are a good place to start, where each bar represents a quantity of something (like units sold) for a specific grouping (like product lines). If there are a limited number of groupings, these types of charts provide a visually

intuitive way to compare quantities, as represented by length or height. Line graphs (see Figure 4.) are also a good choice if the quantities are tracked over time.

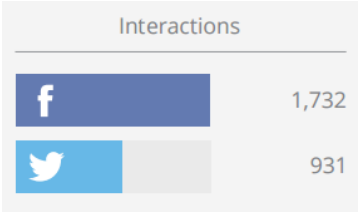

Figure 4. Linear graphs.

Simple bar charts are great if the user needs to rapidly compare different quantities. Otherwise, the interest is focused in the direct relationship between two different quantities; some type of scatter plot visualization would be probably needed (see Figure 5.). When a substantial correlation between the variables exists, a linear pattern will form on the plot. If identifying extreme data points or outliers that do not conform to a normal relationship is important, this could also be a useful way to visualize data for the users.

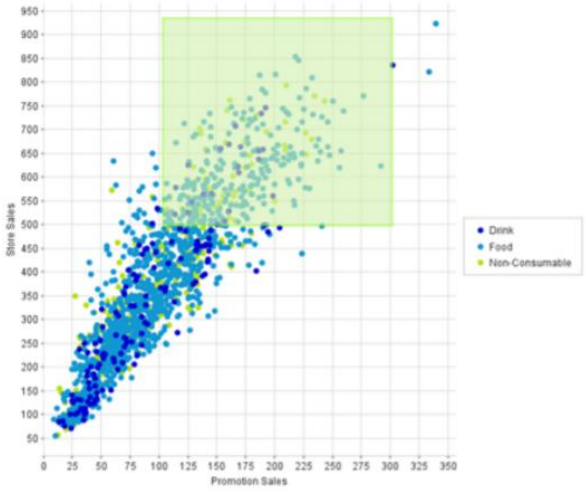

Figure 5. Simple Bar charts.

Scatter plots help show correlation between variables – this example shows the relationship between total store sales (vertical) and sales from a specific promotion (horizontal). The colors represent different product types.

Trends & Changes over time

Time series data is a deep analytics topic, but the crucial point for most visuals it that it adds a new dimension to the analysis, normally where time is displayed on the horizontal axis of your chart (see Figure 6.). The more lines or series you add to a line graph, the more difficult it becomes to understand and interpret at a glance

Figure 6. Basic time series graph plotting sales and gross profit over a month.

These types of charts give users an intuitive feel for changes over time - i.e. growth rates and decline rates (see Figure 7.). When looking at rates of change for a specific measure over time, it can be useful to simply indicate the rate with clear text near the appropriate series on the chart. Showing a directional arrow icon (up, down, or flat) may also provide the user with a clear sense of the change rate.

Figure 7. Displaying a growth rate and icon next to a KPI - in this case entrants in a race.

Shares and proportions

Shares and proportions display a relationship between the parts and the whole, rather than differences over time. Examples here are shares of an investment portfolio allocated to stocks vs. bonds as well as website conversion rates - i.e. the portion of people who opened an email, clicked it, or did not open it. If the primary interest is to display the composition of an aggregate quantity at one point in time - say the breakdown of the investment portfolio by asset class today - then a pie chart would make sense, with slices representing different portions that add to 100% (see Figure 8.).

Figure 8. Basic pie chart showing proportions of medical patients who spent time in different types of hospital rooms.

Pie charts on their own, however, do not convey a tremendous amount of information. If we wanted to show proportions over time, we would use a stacked bar or area chart that can be configured to display total quantities over time broken down by categories in either absolute or percentage terms (see Figure 9.).

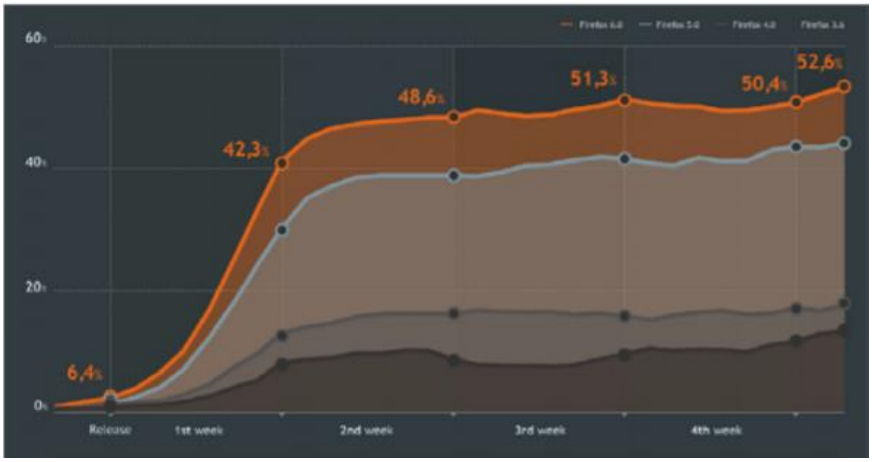

Figure 9. Area chart comparing the uptake over time for different software versions following their initial release.

A unique alternative for showing proportions could also be a radar chart (also known as a spider or web chart), see Figure 10. In this visual, the proximity of the data point to a corner indicates the relative importance of that corner, while several 'webs' can be overlaid to illustrate relative results for different periods or other groupings. Indeed each 'web' creates a different shape representing a different scenario to the viewer.

Figure 10. Radar chart captures the relative importance of different customer churn drivers for a telecom company both in 2013 and in 2012 in a quickly consumable format.

Ranked lists

Presenting a ranked list of the highest or lowest items from a data set is a common dashboard approach (see Figure 11.). Though not really data visualization, this is a good way to provide the 'need to know' information, assuming there is a way for the user to access greater detail. For instance, a sales management dashboard may include quota attainment for the five best and worst performing sales representatives on the team, with the ability to click through to a view that shows all sales representatives with a wider array of metrics for each representative.

RANKING		TRACK	NR STREAMS
	1	Hurry Up We're Dreaming	5,236
	2	Small Things	4,563
	3	Dandelion	4,022
	4	Float	3,785
	5	The Last	3,102
	6	We're Dreaming	4,563
	7	We Stay Together	2,956
	8	One	2,854
	9	Fennesz01	2,820
	10	Passed Me By	2,310

Figure 11. Ranked list of top 10 music tracks by number of streams over a time period.

Geography & location

As you might expect, in this type of visual you are looking at data on a map or physical representation of a real location, which gives the user a richer information experience based on known spatial and physical relationships, a picture is worth a thousand words, as the saying goes.

These locations and their spatial relationships lead to a deeper understanding of behavior and influences, see Figure 12. Since a high percentage of data already have geographic information attached to it, knowledge about these relationships is easily available.

Location intelligence now allows incorporating external data from a variety of sources that can be dynamically combined and updated. Examples of maps with this type of information are heat maps, density maps, thematic or categorized maps and maps with charts (all of them located in one location).

Figure 12. Map of bubbles, each one is in a city, and the volume represents a measure.

Dealing with multiple measurements at once

Usually some of the richest visualizations display more than one type of quantity and one type of grouping in the same space. However, as the variety of information you are trying to communicate increases, so does the risk that your users will miss the point. Two types of multi-dimensional visualizations will be described - bubble-scatter charts and heat grids.

In a bubble-scatter visualization (see Figure 13.), the scatter plot discussed earlier is augmented with one or more additional dimensions. First would be the size of the data point (hence the 'bubble') – where the area of the bubble represents a numeric value associated with that point. This would be followed by color of the data point, to represent another factor.

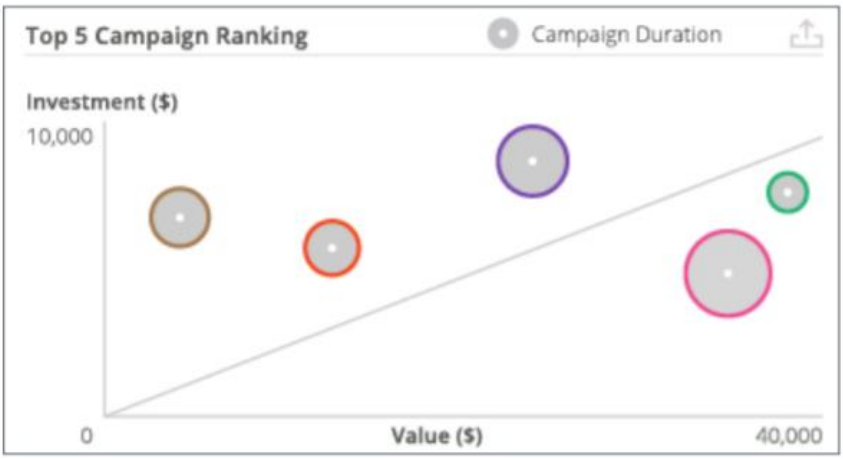

Figure 13. Bubble-scatter chart showing marketing campaigns by value generated (horizontal) and invested cost (vertical), as well as duration the campaign ran for (bubble size).

In a heat grid (see Figure 14.), we can leverage color to convey values or 'temperatures' of data points on a horizontal and vertical access. This type of visual is more appropriate where we expect multiple Y-axis values for every X-axis value, and thus makes sense for representing values sliced and diced by different general or neutral categories. Put another way, a heat grid can help where a stacked bar chart might fall short - rather than trying to display values by size stacked on one another, the size of the data points can be left constant with the colors representing the value scale. That said, the size of a data point could be added in as another dimension in some situations.

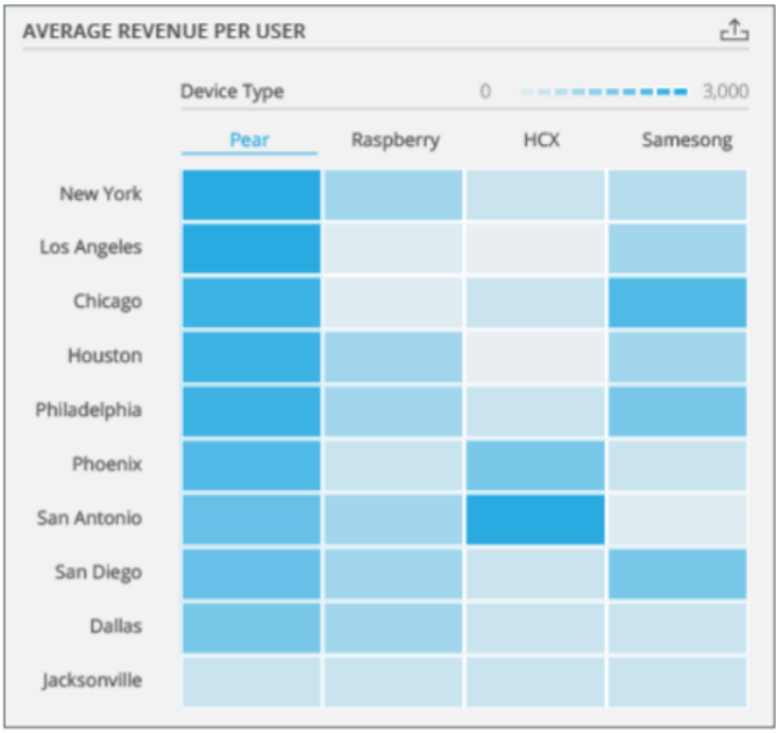

Figure 14. Heat grid displaying average revenue per user (color scheme) across customers, by both geographic market (vertical) and type of mobile device used (horizontal).

Existing main frameworks

This point lists and describes some of the existing main frameworks for building dashboards and data visualization. Main criteria to choose among them are:

- The quality of the solution, which includes usability, learning and dedication time to be able to master it, the options offered, number of graphics, possibilities and the result of visualization.
- Support, community and number of users.
- Priority to free and open source solutions.
- Regarding the purposes, it should be oriented to the web and have attractive, dynamic and interactive graphics.

Therefore, for each solution the more it fits the criteria, the more it will be described. So, some of them, as the pure graphical JavaScript libraries, will only be briefly described, while BI subscribe solutions or SaaS solutions and others will be widely described and compared to each other.

JavaScript libraries to create charts

The following libraries are benchmark when graphing drawings and SVG objects (scalable vector graphics), but they are only going to be mentioned and summarized because most dashboard building frames already use these libraries for graphics. Therefore, they solve much of the work that would have to be done if "native" graphics library would be used. Among the best-known and used are:

RShiny is designed to create analysis output graphs made with R. It is very useful if you have used this language to perform the process of analysis and extraction of data information.

D3.js is an open source JavaScript library for data visualization. It allows you to link arbitrary data to a data document model, and then apply databased transformations to the document. Its emphasis on web standards offers all the capabilities of modern browsers without tying itself to a proprietary framework, while combining powerful visualization components and a DOM-driven data manipulation approach

(Document Object Model). It supports large sets of data and dynamic behaviors of interaction and animation.

This framework follows the imperative paradigm. It offers application help methods that you can use to write code that visualizes your data step by step.

For example, to create a bar chart with D3.js you need to initialize the canvas, calculate where to draw the axis, draw the axis, calculate where to draw the columns, draw columns, legend, point data, add events, etc.

D3 has become a standard for data visualization. There are also libraries built on D3 (dimples, NVD3, xCharts, Rickshaw, C3.js ...) for people looking for top-level APIs.

Chart.js is a small open source library that supports only six types of graphics: line, bar, radar, polar, pie and donut. However, the reason why it is good is that it is sometimes all you need for a project. If the application is large and complex, then libraries like Google Charts and FusionCharts makes sense, while for small hobby projects Chart.js is the perfect solution.

Chart.js is perfect for small projects - when you need flat, clean, elegant, fast JavaScript elements. It is a small open source library taking only 11kb when it is minified and compressed. This includes six main types of graphics (line, bar, radar, polar, pie and donut), each in its own module, so you can only load only what the project needs, making its size even smaller.

It uses the HTML5 canvas element to represent graphics, and polyfills support to work on IE7 / 8. All graphics are responsive.

Frameworks to create dashboards

Freeboard is a free and open source software for dashboard construction. It is not a complete BI solution, but it is easy to integrate with multiple data sources prepared in production, and with an elegant design. You can hire a hosting subscription or install it in your own local environment. The weather data display of the Open Weather Map is already loaded, and it has graphs ready to be fed by JSON data provided by URLs. It is configured to retrieve data from IoT platforms such as Citrix, BuildingLink, Orion, PubNub and Xively. For each data source parameters and data types are specified, as well as update rates. It solves the CORS problem (Cross-origin resource sharing / Cross-domain shared resources).

Additional subscription services: Paid Freeboard subscriptions include free subscription features and provide protected dashboards with private password. You can get five different private boards for $12 a month, 10 for $22 / month, 20 for $42, and 50 for $100 a month. The price of the business subscription must be requested.

Advantages

- It has an intuitive, orderly, well-designed and highly flexible web box.
- Very easy to use and practically no programming skills are needed.
- "Drag and drop" widgets.

Drawbacks: the documentation and error reporting for Freeboard is practically non-existent, so things like incompatibility problems of a data source will not be marked and it will be difficult to find the error and debug a data source that it does not work.

Complete BI SaaS Solutions

Such a solution takes care of both, all regarding preparing the data, and data visualization tools. Preparing the data includes data extraction from different sources like files, operational data bases, services, etc.., prepare and clean it and finally construct the processes for defining the metrics, kpi's and

visualizations. In this section are described the three more popular and used BI software, two non-open source solutions, QlikView and Tableau, both allow you to try their product for free during a trial. If you find one you like and decide to use them for a commercial use, it is expected to pay the big bucks. In addition, one open source platform focused in location intelligence, CARTO.

Tableau is one of the fastest evolving business intelligence and data visualization tool. It is very fast to deploy, easy to learn and very intuitive to use Tableau has a nice user interface and a clean dashboard, which—after you have mastered the learning curve—makes for a good user experience. The simple data drag-and-drop visualization system is the highlight of this software, and many users claim it as the best data visualization tool on the market. These advanced visualization tools allow you to see dots on a map or interesting graphs instead of basic tables. These tools are primarily used in mapping and viewing current trends.

Advantages:

- Can be integrated with R.
- Based on training videos, blog/forum posts, and twitter buzzes Tableau certainly leads the community building effort.

Drawbacks: In order to pull the data, you need out of your system, you will have to Extract, Transform and Load (ETL). This process takes time or puts you in line behind coworkers who have requested other reports to be built. In addition, you can still integrate with many data sources, but the list is shorter than those others are. A private license on Tableau ranges from $999 to $1,999 per user, and gets more expensive depending on server and data access. There is a free desktop version, but keep in mind that the data you use will be made public. Good reasons to look at Tableau if:

- Enticed by really easy, user-friendly drag-and-drop visualization formats.
- Searching for a scalable software solution to use within your organization.
- In need of software, that provides excellent data visualization.

Qlik is a business intelligence data discovery product that is used to create guided analytics applications as well as dashboards designed for business challenges. You can use Qlik Associative Data Indexing Engine to uncover data insights and relationships across various sources.

Qlik exposes data that is not revealed with query-based tools. The tool also offers guided exploration and discovery and collaborative analytics for sharing insights. In addition, the program enables users to create and deploy analytic apps without requiring technical skills. This helps drive quicker response to changing business requirements, shorter time to value, and more insight across a company.

Qlik is more of an all-encompassing dashboard application, while Tableau is more focused on visualization and analytics. It focuses on a variety of business intelligence tools, like Qlik Expressor (a quick and smart metadata intelligence solution) and NPrinting (an application for report generation, scheduling, and distribution). These tools have helped brand Qlik as a powerful reporting solution. Because of its many facets, it is a good solution for enterprise-level companies who can utilize different features across different departments. Other positive features include good third-party integration, advanced data filtering options, and data manipulation.

Prizing: Qlik Personal Edition – offered free with unlimited access. Each private user on Qlik is $1,350 and concurrent users are $15,000. A server license is $35,000. Other services are available at an additional cost.

Good reasons look at Qlik if:

- Part of a large organization that can use its many tools in different departments.

- Looking for a variety of business intelligence tools—-OEM versions, for example—-other than visualization.
- Dealing huge sets of data and need powerful, multifaceted software.

CARTO is a software as a service (SaaS) platform that serves geo-mapping tools to help businesses and designers create web visualizations focused on location. It provides visualizations and dashboards for a comprehensive data analysis tool, an easy-to-use solution that leverages open data and filters, enhancements and analysis to gain deeper insight into BI analysis.

Regarding to the Smart City applications, it is necessary to take into account that much of the data and information to be visualized will refer to geographical points and geographic regions, such as neighborhoods, districts, streets, point locations, etc. In this field, we can conclude that the CARTO platform should be considered a highly adoptable solution to this case.

CARTO Builder is the component that is used to create dashboards and visualizations, see an example in Figure 15. It is a web-based "drag and drop" analysis tool that comes with a variety of public platform datasheets. It is different from its previous versions because of two features: widgets and prediction functions. None of them requires programming knowledge to be used.

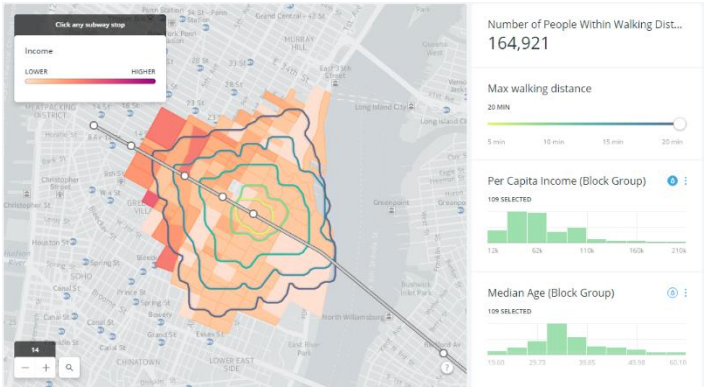

Figure 15. Example of CARTO visualization.

Examples of indicators and variables with localization monitored in a city are:

- Traffic, public transport and mobility in general.
- Contamination and environmental parameters.
- Demography.
- Tourism.
- Others...

Every metric and index can be aggregated in space, like districts and neighborhoods as well as representable on a map.

Features

- Open code.
- Flexible interface, very simple and at the same time advanced, when needed.
- It offers spatial analysis tools such as the ones available in the standard GIS software: overlays, clusters calculations, calculation of service areas by time, proximity, etc.
- No programming experience is needed to create very complete maps or to understand any of the many manuals available.
- For advanced users with programming experience, it offers powerful capabilities through SQL queries and possibility to apply styles to the dashboard using Turbo-Carto, an open-source

CartoCSS preprocessor (pseudo language of cascading web styles), which allows the evaluation of functions in asynchronous mode. Its potential for creativity and data integration is very great.

- In addition to importing numerous file formats, it includes connectors to files hosted in the cloud on platforms like Google Drive and Dropbox. You can also import data from different databases using connectors through its import API.
- It is allowed to synchronize the CARTO tables with the data, either automatically with a maximum frequency of 5 minutes for premium users and 15 for the rest, or manually with ogr2ogr commands. CARTO checks changes in the data, and if so, it proceeds to truncate its table and fill it again completely.

Disadvantages: CARTO platform offers limited services for the non-commercial version, and does not offer privacy for the data in this case. The functionalities can be improved by switching to premium version ($149 per month) and business personal licenses.

Open BI solutions

ELK (ElasticSearch + Logstash* + Kibana) is a solution that is being used to store server logs and obtain metrics and analysis of any type: resource monitoring, log searches, more repeated SQL queries, slower SQL queries, geolocation of users, users that access more times to an application, application usage by users, etc. ELK was designed for centralized monitoring of servers and production applications. It is deployed on a server or cluster in the cloud, and nodes to be monitored and their applications send their logs, which are pre-filtered with Logstash, stored in ElasticSearch and visualized with Kibana to extract metrics and usage patterns in order to improve the performance, see Figure 16. Although this is its initial task, in fact ELK is much more, since it is a solution for data management that allows its transportation between machines, processing and query in real time. Currently there are products based on ELK for business intelligence, which are actually an extension of ELK.

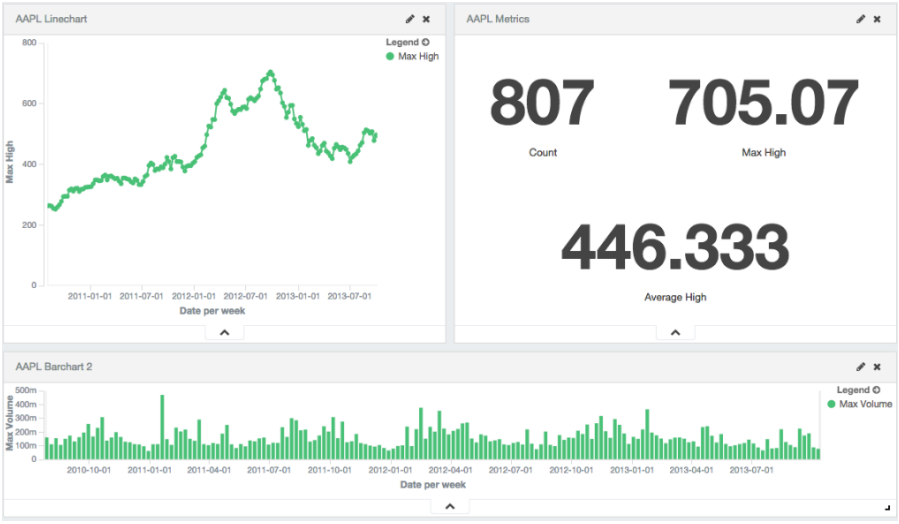

Figure 16. Appearance of Dashboard made with Kibana.

Components

- **Elasticsearch**. Advanced search engine extremely fast. With Elasticsearch, all kinds of data can be searched and filtered through a simple API. The API is RESTful, so it can be used for data analysis, as well for web-based applications production.
- **Logstash**. Tool designed to organize and search log files. However, it can also be used to clean and transmit large data from all kinds of sources in a database. Logstash also has an adapter for Elasticsearch, so these two components interact very well together.

*Although ELK appeared as a joint monitoring solution, the Beats project has recently been incorporated into the group, whereby it is betting that it will be responsible for feeding elastic search, either together or in parallel to Logtash.

- **Beats** is a single-purpose data loader platform (see Figure 17.), which is installed as lightweight agents on its servers to send different types of operating data to libbeat, the Go library (language) containing the common parts of all Beats to handle tasks such as inserting data into Elasticsearch, sending events securely to Logstash, balancing the load of events to multiple Logstash and Elasticsearch nodes, and sending events in synchronous and asynchronous mode. The libbeat platform also includes mechanisms to detect when servers are overloading or the intermediate network is being congested, so it can reduce the sending rate. The beats platform has four official beats: Filebeat, Packetbeat, Metricbeat and Winlogbeat, in addition to these three the development community has created many more, can be consulted from https://www.elastic.co/guide/en/beats/libbeat /current/community-beats.html.

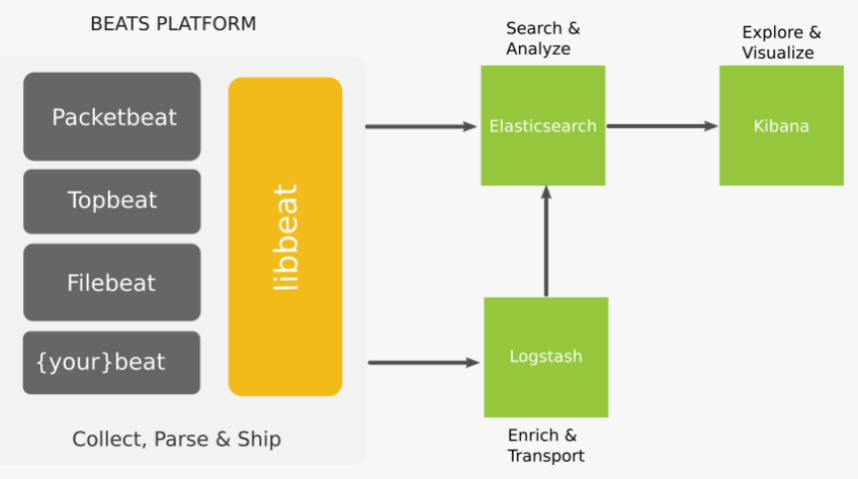

Figure 17. Beats platform.

- **Kibana**. Visual interface for ELK that works in the browser. It is an open source visualization and analysis platform that allows exploring, visualize and represent the data content of an Elasticsearch cluster using the graphics, tables and maps provided by the powerful graphics library d3.

It is good enough to display the data stored in Elasticsearch and does not require programming knowledge, since the visualizations are completely configured through the interface.

Kibana customization

There are two mechanisms for creating custom dashboards. The first is to create a custom control panel from the user interface. The second is to create a custom control panel from a file. Custom control panels can be created from a JSON document template based on a specific schema, or create a custom template as a JavaScript file. By default, it brings examples of dashboards in JSON. There is also a way to create your own visualizations out of the list that Kibana offers, anyway, those cover the most current.

The Eclipse Business Intelligence and Reporting Tools Project (**BIRT**) is an Eclipse top-level project that is focused on tools and capabilities that extract data from data sources, process that data, and present the processed information to the end users. First released in 2004, BIRT is a mature open source-reporting platform with millions of active users.

Actuate, by far the largest code contributor to the project was acquired by OpenText, Canada's largest software company, in January of 2015. At the time of this review, OpenText and Actuate have not indicated any changes to the BIRT Project.

BIRT consists of two main components:

- **BIRT Report Designer**. A graphical report designer that is designed to be easy to use. It is a fat-client tool, installed to the developer's desktop, either within the Eclipse IDE or as a standalone Windows application (the "RCP Designer" package). Report developers use its graphical user interface to create report layouts, hook up data sources, and produce XML-based report designs.
- **BIRT Runtime.** Also known as the "BIRT Report Engine," is a set of Java classes and APIs that takes the XML-based report designs (created by the BIRT Report Designer), "runs" the report designs by querying their data sources, merges the query data into the report layouts, and then produces report output in HTML, PDF, Excel, or other formats. You can embed the BIRT Runtime inside your Java applications. The BIRT Runtime can be deployed either as a set of OSGi jar files, or as a drop-in collection of jar files.

BIRT also includes four additional open-source components that can be used separately from the main components:

- **BIRT Chart Engine**. It is basically just a piece of the BIRT Runtime. With the Charting Engine API (CE API), you can incorporate charting into your Java application without the overhead of the entire BIRT Runtime.
- **BIRT Chart Designer**. OSGi based report designer that can be deployed and used within applications that support the OSGi framework that would like to design reports.
- **BIRT Viewer**. This is a sample "viewer" that is an Eclipse plug-in. It lets you preview reports, populated with the real query data, when you are working within Eclipse. It is also available as a standalone Java EE application. By deploying the BIRT Viewer to a Java application server, you can provide end users with a web-based user interface to view reports, navigate through paginated reports, and render reports into a variety of formats including HTML, PDF, and Microsoft Excel/PowerPoint/Word. The BIRT WebViewer also supports JSP Tag Library extensions for many common report interaction tasks.
- **Eclipse Data Tools Project (DTP)**. BIRT's access to data is designed to flow through the DTP, a stand-alone project. Initially co-sponsored by Actuate and Sybase. The development on DTP has slowed considerably.

The BIRT Report Designer is a robust, powerful tool for creating reports where the output is primarily presented over the web. The BIRT Report Designer is fairly easy to use (although occasionally overly complex), and able to create reports ranging from the simplistic to the very complex with minimal custom code. BIRT has excellent support for essential report features such as charts, crosstabs, using multiple data sources within the same report, re-using queries within reports, adding custom code, and more.

One significant distinction when comparing BIRT to Pentaho and JasperReports is that BIRT lacks an open-source report server. For many applications, a report server provides functionality such as user management, report management, security, notifications, and more. Using a report server often lets report developers concentrate on report creation and the ideal presentation of data, instead of having to code up these separate capabilities. If report server capabilities are important to your application, and you are determined to go the open source route, then you should consider either Pentaho or Jasper. A second significant area where BIRT does not match up to its open source competitors is its incompatibility with OLAP data as describe by either Mondrian or XML for Analysis (XMLA).

Jaspersoft community, JasperReports is one of the most popular and used open source projects for reporting and graphic creation. It has tools both to ETL (Extract, Transform, Load) and to create reports (visualizations). Jasper server has two API, HTML and REST, to interact with it and to ask for its resources. It has a JavaScript library to create dashboards, with the JavaScript API the report is created and embedded in a ready HTML element, without programming; therefore, dashboard layout is 100% configurable, reports are added to the website without further action. The HTML API is used to obtain reports or visualizations, while the REST API is used to list, create, update and delete reports and resources of a Jasper server repository.

JasperReports Server is a standalone and integrative reporting server that allows the delivery of real-time or scheduled information to the web, the printer or a variety of file formats. The iReport and Jaspersoft Studio tools are used to design the reports that will be built and returned in JasperReports Server. Reports can be run or exported to a desired output, or scheduled to run later.

Main features

- Parameters can be sent to the report.
- One report can have more than a graphic, list, etc.
- Jaspersoft is 100% pure Java, so should work in any Java > 1.6 environment.
- It consists of various components: JasperReports server, Jasper reports library, Jaspersoft ETL, visuality.js, IReport and Jaspersoft Studio. (IReport is no longer supported, since it has been replaced by Jaspersoft Studio).
- Application servers: Apache Tomcat, JBoss, IBM WebSphere, Oracle WebLogic, GlassFish and SpringSource tc Server.

Data Sources

- Any RDBMS compliant with JDBC 2.1 and SQL-92 (includes most relational databases).
- Big Data & NoSQL: MongoDB, Hadoop, Cassandra, others.
- Other data sources: Bean, JNDI, XML/A, Custom (e.g. Hibernate, XML, etc.).
- In order to use CSV or Excel files, Jaspersoft ETL or Jaspersoft studio are required for using these files in the Jasper server.

Other features

- It has AWS, professional and enterprise editions.
- Besides support and training, premium services offer extra features.
- Metadata layer. A virtual view of a data source that simplifies presentation in commercial terms while providing data-level security.
- Data virtualization. Allows integration of multiple data sources into a single metadata view to enable analysis and reporting across sources without the need for ETL or a Data Warehouse.

Disadvantages: JasperSoft Studio is a tool somewhat less usable and more complicated than its counterparts in Pentaho and BIRT are. Common operations like creating aggregates is complicated in some cases. This is because JasperReports was born with the main dedication of providing printed reports and later replicate this work for use on the web, so the workflows to create the reports seems a bit more complex compared to the other solutions. One more thing it is the fact the reports must be compiled to be used.

Pentaho is a prominent open source-reporting project with a great vision of a single, comprehensive BI suite that covers the full spectrum of business intelligence (BI) life cycle including ETL, reporting, analytics, visualization, and data mining. Reporting and visualization are only a portion of Pentaho's overall vision of providing a complete open source BI solution. The Pentaho BI Suite Community Edition (BI) encompasses the following open source projects: Pentaho BI Platform & Server, Pentaho

Reporting, Pentaho Analysis Services (Mondrian), Pentaho Data Integration (Kettle), and Pentaho Data Mining (Weka).

Components: The open-source Pentaho BI Suite Community Edition (CE) includes the following components for reporting:

- **Pentaho Report Designer**. A WYSIWIG tool that lets you create reports using a graphical user interface, as opposed to creating reports programmatically or by directly creating and manipulating XML. These reports can then be run by the Pentaho Reporting Engine Classic or the Pentaho BI Server. The Pentaho Report Designer is a stand-alone, desktop-installed client tool, and is not available as an Eclipse or NetBeans plug-in.
- **Pentaho Reporting Engine Classic**. Formerly known as "JFree Report", this is a collection of Java classes and APIs that execute Pentaho's XML-based reports. The Pentaho Reporting Engine runs report designs against data sources, and renders report output in HTML, PDF, Excel, and other output formats. You can embed the Pentaho Reporting Engine inside your Java applications. You do not need the Pentaho Reporting Engine if you use the BI Server.
- **Pentaho Data Integration (Kettle)**. Kettle is a graphical data integration tool that allows developers to build jobs and transactions that can be used to Extract, Transform, and Load (ETL) data from a wide variety of sources.
- **Pentaho Reporting SDK**. The SDK is the Pentaho Reporting Engine Class, plus the documentation and supporting libraries that developers need to embed the Pentaho Engine Classic in their applications.
- **Pentaho BI Server**. The BI Server is a J2EE application that provides an infrastructure for multiple users to run reports and OLAP cubes through a web-based user interface. At the core of the BI Server are the Classic Engine and the Mondrian ROLAP Engine (which run the reports and OLAP cubes respectively), plus a host of server capabilities including authentication, user management, logging, email notification, server APIs, and report scheduling. The BI Server also provides the infrastructure for reports and analytic cubes to access data and metadata via the Pentaho Data Integration's ETL functionality (Kettle).
- **Pentaho User Console**. End-users can login, browse reports, run them, view report results in HTML or PDF, and download report results in other formats. Users can also, for the time being, create basic ad-hoc reports, and conduct some OLAP analysis, however this functionality is likely to be removed in future releases.
- **Pentaho Administrator Console**. Administrators and developers can deploy reports, manage users, set up security access privileges, and deploy workflows.

Pentaho has the editor that is the easiest to learn for creating basic listing reports and for grouped listing reports with aggregations. Its UI was not overly cluttered with sophisticated, less-commonly used functionality, and the tool was easy to learn and performed well. In short, the User Interface is attractive and functionality is sensibly laid out.

Pentaho's stated goal is to provide a comprehensive solution for Data Integration and Business Analytics. This includes solutions for extract, transformation, and load (ETL), basic reporting, data analytics, data exploration, data visualization, and data mining. This broad company vision might explain why the reporting functionality is not as deep as BIRT or Jaspersoft. Functionality that report developers take for granted in BIRT and JasperReports - such as side-by-side report components, cross-tabs, and robust charting - are not as fully developed in Pentaho. Unfortunately, this means when using only Pentaho reporting, it can be more difficult to create complex reports.

A prominent feature is the metadata injection. The ETL Metadata Injection step is able to inject metadata using a transformation template. So, instead of introducing ETL metadata statically into a

dialog box step by step, it is executed at runtime. It is possible to resolve repetitive ETL work loads such as text file upload, data migration and so on.

Conclusions

At the beginning of this work It has been explained that the frameworks that accomplish the criteria would have deeper description in addition they will be compared to each other so, as result of this work it has been made a framework comparison matrix series regarding some concrete key issues that are important to evaluate. It has been analyzed the four-open source and free software, that are BIRT, Jaspersoft, Pentaho and ELK. Table 4.-10. present the topics covered by these matrices and their results.

Table 4. Framework Basic Information.

	BIRT	**Jaspersoft**	**Pentaho**	**ELK**
Open Source Website	eclipse.org/birt	community.jaspersoft.com	reporting.pentaho.com	elastic.co/community
Commercial Web Site	developer.actuate.com	jaspersoft.com/reporting	pentaho.com	elastic.co
License	Eclipse Public License (EPL)	JasperReports Lib LGPLV3	Pentaho Reporting LGPL V2.1	Apache License, version 2 ("ALv2")
Report Designer	BIRT Report Designer 4.4.1	JasperSoft Studio 6.0.1	Pentaho Report Designer 5.2.0-GA	Kibana 5.1
Designer Platforms	Windows, Linux, Mac OS X	Windows, Linux, Mac OS X	Windows, Linux, Mac OS X	Windows, Linux, Mac OS X
Eclipse Plug-in Available	YES	YES	NO	NO
NetBeans Plug-in Available	NO	NO	NO	NO
Standalone Java Client Available	YES	YES	YES	NO
Design Paradigm	web page design, frames, tables and lists	banded reports pixel positioning	banded reports pixel positioning	web page design, frames, tables and lists
Report Compilation	Not required	Required	Not required	Not required
Report Format	XML (.RPTDESIGN)	Report design files (.JRXML) compile to Java Byte Code (.JASPER) Deploy/Run .JASPER files	XML Report file (.PRPT) is a ZIP of Design and other resources	JSON or JavaScript script

Table 5. Report Designer Components.

	BIRT	Jaspersoft	Pentaho	ELK
Common Report Designer Components	Y	Y	Y	Y
Geometric Shapes	Y	Y	Y	Y
Barcodes	Y	Y	Y	Y
Callouts/notes in design mode	Y	Y	N	Y
Table of Contents as standard component	Y	Y	Y	Y
Sub-reports	Y	Y	Y	N
Sid-by-side report components	Y	Y	Y	Y
Tables	Y	Y	N	Y
Cross-tabs	Y	Y	N	Y
Horizontal Panning	Y	N	N	Y
Newspaper/multi-column layout	N	Y	N	Y
Hyperlinks within a report	Y	Y	Y	Y
Actionable charts	Y	Y	Y	Y
Cascading Style Sheets (CSS controlled format)	Y	Y	Y	Y
Conditional Formatting	Y	Y	Y	Y

Table 6. Data Sources.

	BIRT	Jaspersoft	Pentaho	ELK
Multiple data sources and queries per report	Y	via sub-reports/charts	via sub-reports/charts	yes, but no with the wizard
Support for joining multiple data sources in the Designer	Y	N	Y	No due noSql database
Report can further manipulate query data re-sort filter group	Y	Y	Y	Y
Non JDBC Data Sources				
Cassandrea	Y	Y	Y	Y
CSV	Y	Y	Y	Y
Custom Data Adapter	Y	Y	Y	Beats
EJB	Y	Y	Y	N
Excel	Y	Y	Y	Y
Haddop Hive	Y	Y	Y	Y
Hbase	Y	Y	Y	Y
Hibernate	Y	Y	N	N
JNDI	Y	Y	Y	N
JSON	Y	Y	Y	Y
Mondrian	Y	Y	Y	N
Mongo	Y	Y	Y	Y
POJO	Y	Y	Y	N
Script	Y	Y	Y	Y
Web Services	Y	Y	Y	Y
XML	Y	Y	Y	N
XML/ A Server	Y	Y	Y	N
JDBC Drivers	Y	Y	Y	N
Query Designer	Y	Y	Y	Y
Graphical Query Designer	Y	Y	Y	Y
Scripting	JavaScript Java Event	JavaScript Groovy	JavaScript --Bean Script Framework (BSF)	Javascript

	Handlers	Java	-Bean-Script Host (BSH) -Single Value Query -Metadata data-source scripting extension	
RabbitMQ	N	N	N	Y
Logs	N	N	N	Y

Table 7. Output Formats.

	BIRT	Jaspersoft	Pentaho	ELK
Paginated HTML	Y	Y	Y	Y
Unpaginated HTML	Y	Y	Y	Y
XHTML	Y	Y	Y	Y
PDF	Y	Y	Y	Y
Excel (XLS 7 XLSX)	Y	Y	Y	Y
XML	Y	Y	Y	Y
Plain Text	Y	Y	Y	Y
Rich Text (RTF)	Y	Y	Y	Y
Powerpoint (PPT)	Y	Y	N	N
CSV	Y	Y	Y	Y
Postscript	Y	Y	N	Y
OpenOffice report types (doc + sheet)	Y	Y	N	Y
Flash (SWF)	N	Y	N	N
Custom Formats	Y	Y	Y	Y

Table 8. Charts.

	BIRT	Jaspersoft	Pentaho	ELK
Chart wizard	Y	Y	N	Y
Chart Interactivity	mouse-over -tool tips -drill-through -hyperlinks -hide/show series -etc.	hyperlinks	mouse-over -tool tips -drill-through -hyperlinks -hide/show series -etc.	mouse-over -tool tips -drill-through -hyperlinks? -hide/show
Chart themes	Y	Y	N	Y
Precise control over format of all control elements	Y	Y	N	Y
Common Chart Types	Y	Y	Y	Y
Study charts	Y	N	N	Y
Ring chart	Y	N	Y	Y
Tube chart	Y	N	N	N
Cone chart	Y	N	N	N
Pyramid	Y	N	N	N
Time Series	Y	Y	Y	Y
Meter/ Gauge	Y	Y	Y	Y
Waterfall	N	N	Y	N
Step Area	N	N	Y	Y
Step	N	N	Y	Y
Difference	Y	N	Y	Y
Radar / Spider	Y	Y	Y	N
Thermometer	N	Y	Y	N
Candlestic/ Stock Chart (High/Low)	Y	Y	N	N
Bar Sparkline	Y	Y	N	Y
Line Sparkline	Y	N	Y	Y
Pie Sparkline	Y	Y	Y	Y
Maps	N	Y	N	Y
SVG Charts	Y	Y	N	Y

Table 9. Report Parameterization.

	BIRT	Jaspersoft	Pentaho	ELK
Static Parameters select parameter values from a hard-coded list of values	Y	Y	Y	N
Dynamic Parameters users select parameters from a list of values that came from a database	Y	Y	Y	Y
Cascading parameters	Y	Y	Y	Y
Calendar date-picker for parameters of type date.	Y	Y	Y	N
Can specify default values	Y	Y	Y	Yes, but is a plugin or must be developed
Drop-down list boxes	Y	Y	Y	"
Radio buttons	Y	Y	Y	"
Check boxes	Y	Y	Y	"
Combo Boxes	Y	Y	Y	"

Table 10. Aggregations - Summary Data.

Birt	JasperSoft	Pentaho	ELK
Average	Average	Average	Average
Count	Count	Count	Count by Page
Distinct Count	Distinct Count	Count by Page	Extended Stats Ag.
First	Sum	Group Count	Geo Bounds Agg.
Is-Bottom-N	First	Sum	Geo Centroid Agg.
Is-Botton-N-Percent	Minimum	Minimum	Minimum
Is-Top-N	Maximum	Maximum	Maximum
Is-Top-N-Percent	Standard Deviation	Sum Quotient	Percentile
Last	Variance	Sum Quotient Percent	Percent-Rank
Max	System	Calculation	Scripted Metric Agg.
Median	User defined Functions	Count for Page	Stats Agg.
Min		Sum for Page	Sum
Mode		Sum (Running)	Top hits
Moving Ave		Count (Running)	Value Count
Percentile		Group Count (Running)	Bucket Agg.
Percent-Rank		Count Distinct (Running)	
Percent-Sum		Average (Running)	
Quartile		Minimum (Running)	
Rank		Maximum (Running)	
Running Count		Percent of Total (Running)	
Running Sum		User defined Functions	
Standard Deviation			
Sum			
Variance			
Weighted Average			
User defined Functions			

BIRT's continues to provide the strongest report development tool. BIRT's greatest strengths are in its ease of use and the completeness of features. If you are looking for a tool that allows report developers to create reports using a thick client application, which will be deployed into an existing Java application framework, it is hard to beat BIRT. BIRT provides the easiest way to create reports that are focused on delivery over the web.

BIRT has two main weaknesses. First, BIRT is primarily focused on reporting instead of analytics, if you are looking to work with OLAP data, BIRT will not be appropriate. Second, BIRT lacks an open source server component and therefore if you are looking for a complete web-based BI solution BIRT is not the best choice.

Jaspersoft Studio provides an outstanding and widely used report development tool that can easily be deployed either through the Jaspersoft Server community edition or through the JasperReports Library to an existing application. A particular strength of Jaspersoft is the way it works with data passed to the report as plain old java objects (POJOs). Jaspersoft is also the best product if your primary focus is to deliver printed reports.

Jaspersoft's chart engine is significantly weaker than BIRT although we prefer Jaspersoft's implementation of JFreeCharts over Pentaho's. Jaspersoft's Table Of Contents for large report navigation is also more difficult to use than either BIRT or Pentaho. Finally, we found Jaspersoft's SQL editor to be the least developer friendly.

Pentaho Report Designer has particular strengths around its outstanding reporting wizard and the ease of use to create simple to moderate reports. Its ability to connect to and manipulate virtually any data through its Pentaho Data Integration (Kettle) data source is a distinguishing feature. Pentaho's charts implementation and crosstab component (experimental) need to improve to reach parity with the other products. If you are looking for one tool that is reporting-focused, we would choose either BIRT or Jasper over Pentaho.

ELK's advantages include:

- It has a large number of built-in graphics types. The control over them was initially limited, in the version 4 the ability to save items was added, and use "drag and drop" elements to edit dashboards. It is customizable from the web interface, which is the one used for the dashboard itself, so the dashboard creation tool is the dashboard itself.
- The speed and power of Elasticsearch.
- Dashboards are updated in real time as the data changes.
- Very good documentation of all components as well as huge community and support offered by its maintenance team and users.

In the last years, the relational databases have been dominating the way the data is modelled and stored, so the SQL language to retrieve and query data became a standard for these propose. Due to the naturality of a NoSql and no relational database, the different query syntax might be an issue to be learned usually from zero, also it is found that is more complicated to construct advanced query than SQL. However, this is not a drawback itself it is just a matter of time that it turns over the usage of this kind of search engines.

Supplements

To know first-hand about the possibilities of the ELK framework, a different approach of traditional bi software, some tests have been done with some of the most interesting. For them it has started from data of tables of a relational database, data about vessel calls in some ports. The trials done were to

create a simple dashboard with one kpi, a top 10 list and a time series and the raw data showed in a table and being synchronized with the visualizations.

1- For this test it has been exported data from one relational database view to a csv file.

2- Then data will be loaded to ElasticSearch using the Logstash component, to do that there must be a configuration file for the load with the following information and structure:

```
input {
  file {
#Here it is specified, the starting point and sincedb file(this is for posloads or load interruptions, logstash uses it to know until the point is already loaded)
    path => "/vagrant/*.csv"
    start_position => "beginning"
   sincedb_path => "/dev/null"
  }
}
#file type and column names
filter {
  csv {
      separator => ","
     columns => ["COLUMN1", "COLUMN2",... ]
  }
#Here some data is converted
  mutate {convert => ["COLUMN1", "string"]}
  mutate {convert => ["COLUMN2", "integer"]}
  mutate {remove_field => [ "message" ]}
}
# specifying timestamp field/s
  date {
         match => ["FECATRAQUESOL", "dd/MM/YY"]
         target => "FECATRAQUESOL"
        }
#logstash output to elasticsearch
output {
   elasticsearch {
     hosts => "http://localhost:9200"
     index => "vgpestadia"
  }
stdout {}
}
```

Then we tell Logstash to do the load with the command: /opt/logstash/bin/logstash -f /vagrant/confilename.conf.

3- After that, opening Kibana (localhost: 5601) in a browser, we see that there is a new index of name confilename. From here we can build visualizations and the dashboard, see Figure 18.

Figure 18. Pie chart, time series and top ten list visualization build with Kibana.

Kpi, pie chart: Vessel calls by vessel flag (country)

Time Series: Total Gross tonnage of all vessels moored by day, and the line split by the distinct ports.

Top 10 List: Top ten vessels in amount of ports call.

The dashboard can be embedded in a html webpage as an iframe element, also every visualization can be embedded independently too. It is still missing to test how to build visualizations that receive parameters and how to communicate with Kibana. The visualizations are very easy to build, just choose the right one, and with the graphic wizard it is arranged. More complex graphics cannot be built with the wizard, but Kibana has other way to build them that is define the visualization with a json specification. Of course, deep knowledge about how it works is required. Interactions and visualization synchronization. After the dashboard is created, one can click in some part of the graphic, and this triggers to all the other graphics to be synchronized, and then, they are filtered by the value of the field clicked. A tooltip shows up some interesting information about the part of the graphic the mouse hovers on it. All of these features are configurable, and can be enabled or disabled.

Original data can be present and available every time. You can create a search where all data is included, and the filter is updated when interacting with the graphs, so we see the composition of this search, in addition the table can be sorted by fields and is paginated (no need to handle huge amount of records do not reduces Kibana performance), so at all times have the complete dataset that make up the visualizations.

3.6 Multi-modal Planning

Finding computing best possible routes in road networks from a given source to a given target location is a frequently deal for many people. There are also many applications like logistic planning or traffic simulation that need to solve a huge number of such route queries. Route planning techniques have evolved quickly in terms of efficiency and accuracy during last years. Two of the first routing algorithms proposed consider the distance (Dijkstra, Bellman-Ford, using a predefined metric) between one node to

each other in a graph. The Classic Algorithm A* extends these calculations with supplementary information and a greedy strategy to improve its results. Meta-heuristic methods have been also used successfully in the computation of the shortest path. Nevertheless, these solutions yields very slow query times when we deal with realistic road networks, this fact disable its use in real-time or interactive applications (Schultes 2008; Geisberger 2011). On the other hand, applying aggressive heuristics not always achieve accurate results. Road networks present structural properties (e.g. networks are sparse, layered, almost planar and present hierarchical structures) that support different speed-upper techniques. These techniques are based on a preprocessing step, where auxiliary information about the network is obtained and annotated, that are used to accelerate all subsequent queries. Speed-up techniques can be classified on: Goal-Directed Search techniques, different algorithms can be found: geometric A* search, heuristic A* search based on distance estimations, landmark-based A* search, signposts, geometric containers or edge flags, managing different graph annotation schemes and exploration and Hierarchical Approaches, that takes advantage of the topological properties of the network (e.g. the separator-based multi-level method or the popular contraction hierarchies) (Brodesser 2013).

From the side of operation research, some relevant routing problems and variants have been labeled; for some, specific solution methods exist, or general approaches as: metaheuristics or graph exploring methods can be applied. Two recent reviews are (Braekers, Ramaekers and Nieuwenhuyse 2015; Gavalas et al. 2016). Some applications on the tourism domain are the tourist trip design problem (TTDP) refers to a route-planning problem for tourists interested in visiting multiple points of interest (POIs). Proposing user touristic routes according its preferences, several mobile applications just provide route recommendations (e.g. Google City) and specific algorithms have been developed to satisfy different variant of the problem.

Within the real-time graphic information systems exist different solutions and implemented, among which we can find the services of planning and tracking routes. Bellow, some relevant examples and the summary of their content in Table 11.:

- **Google Maps** is a web map application server belonging to Alphabet Inc. It offers scrollable map images as well as satellite photos of the world and even the route between different locations or street images with Google Street View. (https://www.google.es/maps/)
- **RouteRank,** route calculator. It accumulates "all the available information" of the European means of transport and of many airports around the world. With these data, the system theoretically calculates the most efficient transport / route combination in terms of CO_2 emissions. The possibilities include travel by car, plane, train or combination of two or more of these. Additionally it provides information of prices and cost of the trip, and allows to directly acquire the tickets (by plane or train).(https://www.routerank.com/fr/)
- **Rome2rio** is a search engine that searches for any city, town, monument, attraction or address worldwide. It finds thousands of multimodal routes to take you simply from A to B. It contains all the information you need to compare and choose the best transport options for your trip. Search for flight, train, bus, ferry, rideshare and rental car information including prices, journey duration and booking details. (https://www.rome2rio.com/es/)
- **Vivanoda** compares and combines air, bus, rail, ferries and car sharing all in one search. (http://www.vivanoda.co.uk/)
- **Waymate** is an online platform that allows intercity travelers to find, compare, and book available modes of transport between their origins and destinations in a reliable, fast and easy way, by accumulating multiple modes of transport and reporting real-time information. It compares all available travel options. (http://waymate.de)
- **Wanderio** is a web application that takes the online travel experience to the next level, taking care of the user from doorstep to final destination. Wanderio compares flights, trains, ferries and

ground transportation services, to let the user choose and book the solution that suits him best, sorting alternatives by price, travel time, and CO_2 emissions. (http://wanderio.com)

- **Moovit.** Moovit offers real-time public transit information and GPS navigation across public transport modes, including buses, ferries, metro/subways/undergrounds, trains and trams, among others. Users can plan trips across transportation modes based on real-time data, integrating official public transit data from transit operators with real-time data collected from users via crowdsourcing. (http://moovitapp.com)
- **TripHobo**, the ambition for TripHobo is to be the central point of the trip planning process by allowing users to plan, optimize, and eventually book customized trips by leveraging technology. (https://www.triphobo.com/create-your-itinerary)
- **Expedia** is an Internet travel agency and has its headquarters in the United States with delegations in 31 countries. It book plane tickets, hotel, car rental, cruises, vacation packages and various theme parks through the web and phone. The website (https://www.expedia.es/) uses multiple global distribution systems to make reservations such as the Saber system for booking flights and hotels, Worldspan, and Pegasus in addition to its own reservation system of hotels belonging to its network.
- **fromAtoB**, is a multimodal journey planner. It is a search engine for cheap train offers, flights, ride shares and bus connections (https://www.fromatob.com/). It can search for a connection between two destinations within Europe. In the process making, all relevant means of transportation such as train, airplane, bus, and ride share are compared and combined with another, independently of their brand.

Table 11. Content of different applications.

Travel Portal	Flight	Ferry	Car	Train	Bus
Google Maps	Yes		Yes	Yes	Yes
RouteRank	Yes		Yes	Yes	
Rome2Rio	Yes	Yes	Yes	Yes	
Vivanoda	Yes	Yes		Yes	Yes
Waymate	Yes		Yes	Yes	Yes
Wanderio	Yes	Yes		Yes	Yes
Moovit	Yes	Yes		Yes	Yes
TripHobo	Yes	Yes	Yes		Yes
Expedia	Yes	Yes	Yes		
FromAtoB	Yes		Yes	Yes	Yes

Specifically in the context of alternative transport modes and Mobility-as-a-Service (MaaS), some companies as SocialCar (https://www.socialcar.com), European Travellers Clubs (www.europeantravellersclub.eu/), TAP TSI (*tap-tsi.uic.org/*), Smart Ticketing Alliance (www.smart-ticketing.org/) and others, are providing data to multi-modal planning services.

From the perspective of developers, some software is available for creating new routing applications/deployments mainly supported by OSM data. **OpenTripPlanner (OTP)** (http://www.opentripplanner.org/) is an open source platform for multi-modal and multi-agency journey planning that provides a range of passenger information and transportation network analysis features using our infrastructure for finding itineraries combining transit, pedestrian, bike, and car segments. It follows a client-server model, providing several map-based web interfaces as well as a REST API for use by third-party applications. OTP relies on open data standards including GTFS for transit and

OpenStreetMap for street networks. Launched in 2009, the project has attracted a thriving community of users and developers, receiving support from public agencies, startups, and transportation consultancies alike. Other open-source tools are OSRM (http://project-osrm.org/), VALHALLA (https://github.com/valhalla) or OpenRouteService (https://openrouteservice.org/).

In the orbit of European R&D Projects, some relevant projects running under FP7 /H2020 including projects selected under the FP7-SMARTCITIES-2013 call as well as other projects under CIP ICT-PSP and Transport program. Some projects have work on multi-modal planning services: MoveUs (www.moveusproject.eu), MyWay (www.*myway-project.eu*), Streetlife (*www.streetlife-project.eu/*), PETRA (http://petraproject.eu/), OPTICities (*www.opticities.com*), MOBINET (*www.mobinet.eu/*), MobilWallet (http://www.mobiwallet-project.eu/), BonVoyage (*bonvoyage2020.eu/*), MASAI and many other initiatives.

Most of the projects agree that three main enables are multi-modal trip planning, user profiling (or patterns) and on-trip user support services. Here, some details of the most relevant.

- **Moveus Project:** ICT Cloud-Based Platform and Cooperative Systems for Personal Mobility, Available, Universal and Safe. Provides best multimodal journey options in terms of different objectives (times, energy efficiency, incentives, etc.) according to user preferences and the real time/expected situation.
- **eCOMPASS:** eco-friendly urban multi-modal route Planning Services for mobile users. It introduces new mobility concepts and establishes a methodological framework for route planning optimization following a holistic approach in addressing the environmental impact of urban mobility. It has been developed a set of tools and services for end users to enable eco-awareness in urban multi-modal transportations. The eCOMPASS project presents an updated deep analysis of current problems, algorithms and underlying approaches. Basically, there are two different approaches to cover timetable requirements: time-expended and time-dependent, each one, prioritizing flexibility under changing constraints or memory needs and response time. (http://www.ecompass-project.eu/)
- **MyWay** is a FP7 project started in September 2013 and ending in February 2016 with a Consortium of 14 members. The project aims at developing an integrated ICT service platform, the European Smart Mobility Resource Manager, developed as an open, modular and scalable framework able to seamlessly integrate different types of urban transport modes by enhancing existing Trip Planning and Travel Demand management services. An impact is expected in terms of reduction of the traffic congestion thanks to improvement in mean journey times and significant shift in travel choices from private to collective (Public transport) and flexible (flexible transport, shared e-scooters, bikes) modes. MyWay operates in several directions to reach the objectives. These include:
    - Improvement of service offering, accessibility and interoperability.
    - Improvement of development conditions for the providers with the reduction of market entry barriers.
    - Application of ICT technologies for smart mobility.
    - Development of multimodal journey planning systems based on awareness of the resources.
    - Flexible and intelligent mechanisms applied to Trip Planning like personalized discovery, negotiation-based matchmaking.
- **MOVESMART**. Started in November 2013 and with duration of 36 months the Movesmart consortium includes 11 partners. The project will be validated in the two pilot cities of Pula-Pola in Croatia and Vitoria-Gasteiz in Spain. The project aims at providing a route planner featuring a time-dependent route determined with a set of crowd-sourcing tools that support the collection of

real-time information by travellers. The crowd-sourced information is used to improve the services (for example in terms of traffic prediction) and to provide context-aware journey planning capabilities. In Vitoria-Gasteiz a special use case is related to efficiently plan route for FEV in order to maximize the operational range of the vehicles and to offer the most eco-friendly route (reduction of the CO_2 emissions).

- **PETRA Project.** Personal Transport Advisor: an integrated platform of mobility patterns for Smart Cities to enable demand-adaptive transportation system. Petra is founded under the SmartCities call. It started in February 2014 with a total duration of 36 months. The consortium includes seven partners from local authorities, enterprises and research institutes. The main objective of PETRA is to develop a service platform that connects the providers, the controllers and the travelers in a City. The travers are expected to contribute on solving the trip planning problem by providing better average travel time and information on better integration of private and public mobility schemes as well as on re-planning capabilities under uncertainty with some specific use cases like drive-park-ride.

3.7 Adaptive control

Adaptive control (AC) is defined as the capability of the system to modify its own operation to achieve the best possible mode of operation for highly variable and uncertain conditions without a priori information to represent the system. This requires a control method in which one or more parameters are sensed and used to vary the feedback control signals in order to satisfy certain performance criteria. A well-known example is coming from the aircraft industry, in which AC is used to modify the operation mode depending on the weight of the plane, which slowly decreases due to fuel consumption.

AC is founded in parameter estimation. Recursive least squares and gradient descent are two methods to realize this and are commonly used to modify estimates in real-time. The more advanced algorithms for AC can be distinguished along several axes (e.g., feedforward vs. feedback AC, direct vs. indirect), and an increasing amount of these are integrating machine learning techniques to optimize and learn the control function (e.g., the use of Markov decision processes and Support Vector Regression for resource allocation in uncertain and changing environments (Csaji and Monostori 2008). More recently, machine-learning techniques have also been used for adaptive control of smart environments. For example, Stenudd (2010) provides a model for using machine learning for adaptive control in smart environments (e.g., to perform resource allocation in a distributed multimedia streaming application). Tao (2014) provides recent survey of the fundamental theoretical aspects and technical issues of multivariable adaptive control, as well as an overview of current approaches.

3.8 Cloud Management Platforms for IoT

Virtualizing sensor nodes and managing them through a uniform Cloud management platform is addressed by some general research efforts and some open source and commercial solutions. For billions of devices to sense, communicate, compute, and actuate, traditional data management approaches need however to be enhanced with new developments and standards.

Relevant European projects are IoT-A (http://www.iot-a.eu/public), which studies IoT architectures and FI-WARE (https://www.fiware.org/) in which tools are developed for the creation of new applications and Internet services. FI-WARE is currently extended with enablers to build IoT custom platforms.

Kaa (http://www.kaaproject.org/overview/) is a hardware-agnostic middleware platform for building complete end-to-end IoT solutions, connected applications, and smart products. Other platforms, such as Smart M3 (http://sourceforge.net/projects/smart-m3) and HyperCAT (http://www.hypercat.io/), focus

on semantic interoperability among devices. Different research initiatives are performed in which the process of sharing of sensors and actuators as well as their data through cloud technology. Some models are presented by Distefano et al. (2013) and Ibbotson et al. (2010). The COMPOSE (http://www.compose-project.eu/about) project proposes an open and scalable marketplace in which smart objects are associated to services. Solutions for cloud-based sensor data acquisition and management platforms that could be reused in the project are Nimbits, Paho, ThinkSpeak, IBM Mote Runner and Sensor Cloud. It will also be possible to use sensor frameworks based on runtime software on the device, which provides a unified sensing interface towards applications. A specific sensor driver will transform sensed data in data to be used in the unified interface. Next to this existing sensor, frameworks like the ODK Sensor Framework and the Sensor Data Collection Framework (SDCF) can be used.

3.9 Big Data processing

The processing of very big amounts of information (in the order of Terabytes and even Petabytes) has been an area of great interest during the last years and some remarkable breakthroughs have been achieved to the point of giving birth to the so-called Big Data movement. The foundational key concept in which Big Data is based is the functional programming inspired MapReduce (Dean and Ghemawat 2008) algorithm. Having this processing model in its center the highly influential Hadoop (https://hadoop.apache.org/) framework was developed and served as a basis to give birth to a new data based industry. This framework still serves as the base for the processing of the vast majority of high volume data processing systems in many real life applications.

During the first stages of this new Big Data industry the main attention has been focused on the processing of large amounts of data in a batch way, i.e., typically data processing has been organized in several steps, starting with data acquisition, storing, and finishing with the processing of all this already stored information. However, this was only the first step. New applications require real-time processing of information, especially in the field of IoT, where data will be continuously been generated by an always increasing number of sensors and autonomous devices. Some new frameworks have been created during the last years, both tackling the real- time processing of streams of information, such as Storm (https://storm.apache.org/) or even also highly improving performance of processing functions and implementing additional data processing models besides Map Reduce, as made by Apache Spark (http://spark.apache.org/streaming/).

Big-data batch processing

Big-data batch processing is a MapReduce-based parallel computing paradigm of cloud computing. There are several tools and techniques are based on batch processing and run on top of Hadoop. These include Mahout, Pentaho, Skytree, Karmasphere, Datameer, Cloudera, Apache Hive, and Google Tenzing.

Mahout (http://mahout.apache.org/) was introduced by Apache and takes a scalable, parallel approach to mining big-data. It is used in large-scale data-analysis applications. Google, IBM, Amazon, Facebook, and Yahoo have all used Mahout in their projects. Mahout uses clustering analysis, pattern analysis, dimension reduction, classification, and regression.

Skytree (http://www.skytree.net/) is a general-purpose server with machine learning and advanced analytics for processing huge datasets at high speed. It has easy commands for users. Machine learning tasks in Skytree server include anomaly or outlier detections, clustering analysis, regression, classification, dimensions reductions, density estimation, and similarity search. Because its main focus

is real-time analytics, it enables optimized implementation of machine-learning tasks on both structured and unstructured big data.

Pentaho (http://www.pentaho.com/product/big-data-analytics) is a big-data software platform for generating business reports. It is enables data capturing, integration, exploration, and visualization for business users. With business analytics, the user can make data-based decisions and increase profitability. Pentaho uses Hadoop for data storage and management and provides a set of plugins to communicate with a document-oriented model of NoSQL databases (i.e., MongoDB) and Cassandra database.

Karmasphere (http://www.karmasphere.com/) is a platform for business big-data analysis. It is based on Hadoop. With Karmasphere, a program can be efficiently designed for big-data analytics and self-service access. Karmasphere is capable of big-data ingestion, reporting, visualization, and iterative analysis in order to gain business insight. It can process structured and unstructured big data on Hadoop embedded with Hive.

Datameer (http://www.datameer.com/) provides a business integration PaaS, called Datameer Analytic Solution (DAS), which is based on Hadoop and is used to analyze a large volume of business data. DAS includes an analytics engine, data source integration, and data visualization (in the form of reports, dashboards, and charts). DAS services are deployed in other Hadoop distributions, such as Cloudera, Yahoo!, Amazon, IBM BigInsights, MR, and GreenplumHD. Because the main objective of Datameer is data integration, data can be imported from structured data sources, such as Oracle, MySQL, IBM, HBase, and Cassandra, as well as from unstructured sources, such as log files, LinkedIn, Twitter, and Facebook.

Cloudera (http://www.cloudera.com/) provides Hadoop solutions such as batch processing, interactive search, and interactive SQL. Cloudera is an Apache Hadoop distribution system called CDH that supports MR, Pig, Flume, and Hive. Cloudera also supports embedded plugins with Teradata, Oracle, and Neteza.

Real-time-based or stream-based processing

Stream-based processing techniques are used to compute continuous flows of data (data streams). Real-time processing overcomes the limitations of batch-based processing. Projects that use stream processing include Storm, S4, SQLStream, Splunk, Kafka, SAP Hana, Infochimps, and BigInsights.

Storm (http://www.stormproject.net) is a fault-tolerant, scalable, distributed system that provides an open-source and real-time computation environment. In contrast to batch processing, Storm reliably processes unbounded and limitless streaming data in real-time. Real-time analytics, online machine learning, interactive operating system, and distributed remote procedure call (RPC) are all implemented in Storm project. This project allows the programmer to create and operate an easy setup and process more than a million of tuples per second. Storm comprises different topologies for different Storm tasks created and submitted by a programmer in any programming language. Because Storm works through graph-based computation, it has nodes, i.e., spouts and bolts, in the topology. Each of these nodes contains a processing logic and processes in parallel. A source of streams is called a spout, and a bolt computes input and output streams. A Storm cluster system is managed by Apache ZooKeeper.

In 2010, Yahoo! introduced S4 (Neumeyer 2010), and Apache included it as an Incubator project in 2011. S4 is a platform that facilitates fault-tolerant, distributed, pluggable, scalable computing. It is designed to process large-scale continuous streams of data. Because its core library is written in Java, a programmer can easily develop applications in S4, which supports cluster management and is robust, scalable, and decentralized. It is used to process large-scale data streams. Analogous to Storm, S4 can

also manage the cluster by using Apache ZooKeeper. Yahoo! has deployed S4 for computing thousands of search queries.

SQLStream (http://www.sqlstream.com/blaze/s-server/) is a platform for processing large-scale unbound streaming data in real-time with the support of automatic, intelligent operations. Specifically, SQLStream is used to discover interesting patterns in unstructured data. The platform responds quite rapidly because the streaming data is processed in memory. Server 3.0 is a recently released version of SQLStream and is used for real-time big-data analytics and management.

Splunk (https://www.splunkstorm.com/) is a platform for analyzing real-time streams of machine-generated big data. Senthub, Amazon, and Heroku have all used a Splunk big-data intelligent platform to monitor and analyze their data via a web interface. Splunk can be used with structured or unstructured machine-generated log files.

Kafka (Auradkar 2012) has been developed for LinkedIn. Kafka is a stream processing tool for managing large-scale streaming and messaging data and processing it using in-memory techniques. Kafka generates an ad hoc solution to the problems created by two different types of data, i.e., operational and activity, belonging to a website. Service logs, CPU/IO usage, and request times are examples of operational data that describes the performance of servers. Activity data, on the other hand, describes the actions of different online users' actions. These actions include clicking a list, scrolling through webpage content, searching keywords, or copying content. Kafka is used in several organizations.

SAP Hana (Kraft 2012) is a stream processing tool that also processes streaming data in-memory. SAP Hana is used for real-time business processes, sentiment data processing, and predictive analysis. It provides three real-time analytics: operational reporting, predictive and text analysis, and data warehousing. SAP Hana can also work with interactive demographic applications and social media.

Infochimps (http://www.infochimps.com) cloud suite covers several cloud IaaS services, categorized as:

- cloud streams: real time analytics for multiple data sources,
- cloud queries: query capability for NewSQL and NoSQL (i.e., Apache Cassandra, HBase, MySQL, and MongoDB)
- cloud Hadoop: analysis of massive amount of data in HDFS. Infochimps platform is suitable for both private and public clouds. It can also control STORM, Kafka, Pig, and Hive.

BigInsights (http://www-01.ibm.com/software/data/infosphere/biginsights/) is used in the Infosphere platform introduced by IBM. BigInsights manages and integrates information within Hadoop environment for big-data analytics. BigInsights leverages InfosphereStreams, a stream-based tool of the IBM Infosphere. BigInsights is used for real-time analytics on large-scale data streams. JAQL, Pig, Hive (for querying), Apache Lucene (for text mining), and Apache Oozie (job orchestration) are supported by BigInsights.

Big-data Research Directions and Challenges

High complexity of applications and the evolution of the data market recently require not only batch and real-time streaming processing, but also their integration. In other words, results provided by historical data analysis need to be used to analyze streaming data and vice versa. To achieve this goal some remarkable architectural models such as Lambda architecture (http://lambda-architecture.net/) (a data-processing architecture designed to handle massive quantities of data by taking advantage of both batch - and stream-processing methods) have been proposed (Marz and Warren 2013). In addition, currently cutting edge projects trying to locate streaming in the center of data processing are in active development phase, such as Flink (https://flink.apache.org/) which is intended to serve as the next step

in Big Data frameworks evolution or Apex (http://apex.incubator.apache.org) (and its commercial version, DataTorrent (https://www.datatorrent.com/)) which tries to unify batch and stream processing under the same model.

The algorithms for understanding the data within Big Data are often called data analytics (Delen and Demirkan 2013). Within the field of data analytics, many people recognize a separation in descriptive analytics: describing what the data looks like; predictive analytics: predicting what is going to happen; and prescriptive analytics: predicting what should happen to reach a goal. In 2009, ETSI started a Technical Committee on M2M communication, which is aimed at defining a common architecture to achieve a fully horizontal view, where a common network and application infrastructure can connect different M2M domains and can be shared by different applications. Nowadays existing M2M solutions have been defined from the device or the service point of view, disregarding the exchanged content. This is one of the failures of current IoT: there is a split between things and the content they produce and communicate. BETaaS (Building the Environment of Things as a Service) project made progress in this direction, but it has been applied to Smart City and Home Automation use cases, with the general public in mind.

3.10 Mobility pattern analysis; spatial big data

Spatial data mining is the method of discovering interesting and earlier unknown, but potentially useful patterns from the spatial and spatiotemporal data. However, explosive growth in the spatial and spatiotemporal data and the emergence of social media and location sensing technologies emphasize the need for developing new and computationally efficient methods tailored for analyzing big data.

SBD are important to society, they are used for Eco-routing, Public Safety & Security and understanding Climate Change. SBD exceed capacity of current computing systems in order to manage, process, or analyze the data with reasonable effort and due to Volume, Velocity and Variety. The main difference between data mining in relational DBS and in spatial DBS is that attributes of the neighbors of some object of interest may have an influence on the object and therefore have to be considered as well. The explicit location and extension of spatial objects define implicit relations of spatial neighborhood (such as topological, distance and direction relations) which are used by spatial data mining algorithms. Therefore, new techniques are required for effective and efficient data mining.

Different Spatial patterns are relevant in this field: spatial outlier, discontinuities: bad traffic sensors on highways, location prediction models: model to identify habitat of endangered species, spatial clusters (e.g. in other domains, crime hot spots, cancer clusters and co-location patterns).

Specifically, mobility is based on the movement of people, the quality or state of being mobile, the ability to move physically, movement within or between social classes and occupations. Mobility patterns allows for prediction using certain metrics such as velocity, current location, direction, etc. Factors such as affordability, travel time, travel cost, convenience, flexibility, technology, and its relative level of service are used to predict the travel demand. One way to perform mobility prediction is to extract the regular motion patterns from mobility traces. This allows one to compare the current trajectory of a given mobile to those patterns, and then infer the end of the trajectory; extracting regular patterns from noisy traces very naturally leads to artificial intelligence oriented techniques.

For example, Mobility Spatiotemporal patterns are introduced by a data-driven framework that predicts the user's next places using his/her past visiting patterns analyzed in the context of mobile device logs. Specifically, the notion of the spatiotemporal-periodic (STP) pattern is proposed to capture the visits with spatiotemporal periodicity by focusing on a detail level of location for each individual.

The extraction of data Mobility patterns can use different data sources (Anda, Fourie and Erath 2016), being the most relevant and promising the following:

- **GPS data for mobility patterns**. A typical GPS trajectory is a trace is a time-stamped sequence of pairs of latitude and longitude. Many GPS devices also record the speed, altitude, heading, bearing and other generic information related to the mobile device. Each datum is labeled by a key that identifies each device. Because of measurement errors, the GPS coordinates of the same location may vary from time to time, being needed the process of mapping the raw readings to discrete location symbols.
- **Mobile phone trace**. While conventional mobile phones usually only sporadically exchange information with cell-towers, smart phones continuously provide information. Given this fact together to its high level of penetration rate, constitutes a promising base for travel demand modelling. The most relevant characteristics are the following:
    - Smart phones not only exchange data much more frequently with the mobile network provider allowing more continuous tracking.
    - Mobile phone networks, whether it is GSM, CDMA or LTE, mobile phone networks require regular and frequent handshakes with the cellular communication antennas.
    - Constantly and frequently determining the location. The user's location is calculated by determining the location of the cell antenna closest to the handset.

    The information exchange includes two types of events:
    - Network-triggered location updates occur when a mobile phone is switched on and connects to the cellular network, moves between two different cell areas (i.e. handover) in a call, moves to a new LA or polled by the network as its associated timer ends.
    - Event-triggered updates happen in the following situations: a call or SMS is placed/sent or received or when the user connects to the Internet (e.g. email/Watshup periodically server check)

- **Smart Card Automatic**. Fare Collection or transactional systems are using any identification and transaction protocol (e.g. radio-frequency identification (RFID) technology, magnetic band, TSC technologies, etc.). For the case of public transportation, users check their entrance and exit, so additionally to revenue collection; large quantities of individual detailed information are collected (e.g. boarding times, boarding stations, alighting times, alighting stations and vehicle identification). This represents a huge potential in better understanding travel behavior and improving current transport systems.
- **Social Networks.** A Social Network is defined as a set of persons who are linked pairwise, so that each person can reach any other through an active tie [Axhausen, 2006]. It is argued that the geography of the members of the travelers' social networks together with the geography of the relevant activity locations determines the amount and style of travel. Social network data is being collected mainly in person, using ego approaches and combining qualitative or quantitative techniques. Some relevant researches around social networks exploitation, with application to different problems are (spatial location, frequency of interactions, etc.) (Ohnmacht et al. 2015; Carrasco et al. 2008; Van den Berg, Arentze and Timmermans 2009; Van den Berg and Timmermans 2015) and information on habitual mobility, preferences, biographical questions (Axhausen and Frei 2007).
- **Surveys.** Provide datasets related to very different areas. Such datasets can be used to validate home and working areas, city patterns such as hotspots, commuting, traffic, flows, and land use. Other use of Census information is that it provides the means to perform scaling expansion from information derived by large-scale human mobility sensors. The main advantage is the very

refined spatial resolution that is often the census block. The main disadvantages are that they are updated usually only every 5 to 10 years.

The following list provides a classification of the algorithms with respect to their functionality:

- **Cleansing Algorithms:** The data captured may have noise or be corrupted due to errors in the acquisition process. Algorithms such as Map-Matching are responsible for the adjustment of the data captured by the GPS sensors, to get valid positioning data of the users in the road network. Concerning map-matching algorithms, we can find different algorithms approaches (Olabarrieta, Torre and Molinete 2014): Geometric algorithms only consider the position obtained by the GPS and road geometry. Different approaches have been proposed: topological-based (Greenfeld 2002), probabilistic algorithm (Ochieng, Quddud and Noland 2003), other algorithms include the use of Kalman filters (Ochieng, Quddud and Noland 2003), particle filters (Gustafsson et al. 2002), bayesian methods (Pyo, Shin and Sung 2001), fuzzy logic (Quddus, Ochieng and Noland 2007), among others.
- **Analyzing Algorithms:** these algorithms are responsible for the transformation of the original data into information of the current state of the road, resulting in traffic variables like flow, travel times and others. In general, fusion operators can be classified according the information managed. Soriguera (Soriguera 2011) proposes the following taxonomy: Context Independent Constant Behavior Operators (CICB), where only measures are taken into account (e.g. Bayesian and Dempster-shafer methods), Context Independent Variable Behavior Operators (CIVB), where the behavior is function of the variables to fuse (e.g. expert systems) and Context Dependent Operators (CD, where the operation depends also of the global knowledge (e.g. linguistic-based fuzzy-logic). Independently of the specific methods and sources, data fusion techniques allow fusing different inputs, obtaining more reliable estimations than originals (Cipriani 2012; Koch 2010).
    - **Mobility Pattern Analysis.** Different spatial patterns are relevant in this field: spatial outlier, discontinuities: bad traffic sensors on highways (DOT) (Buchin et al. 2010; Buchin et al. 2011), location prediction models: model to identify common POI or usual trajectories, spatial clusters (e.g. in other domains, crime hot-spots (NIJ)) co-location patterns (Buchin, Kruckenberg and Kölzsch 2012; Järv et al. 2012), periodic and frequent patters and clustering (Demšar and Virrantaus 2010; Laube and Purves 2011; Mirkovic et al. 2011). The management of such information needs specific technical capabilities (Pelekis and Theodoridis 2014; Renso, Spaccapietra and Zimányi 2013).
    - **Demand model estimation.** In Dynamic Allocation Models Traffic - Dynamic Traffic Assignment Models (DTA), the problem of merging data from multiple sensors predispositions has been analyzed in the literature of traffic engineering. Ashok and Ben Akiva (Ashok and Ben-Akiva 1993; Ashok and ben-Akiva 2000) formulated an approach to the problem of real-time estimation and prediction of OD matrix based on a state space model, determined using a Kalman filtering algorithm. The use of deviations from the OD flows Matrix historical values provides an elegant framework for incorporating structural information (generated during calibration system offline) in the on-line process. The approach has been applied in the Dynamit DTA system (Antoniou et al. 1997; Ben-Akiva et al. 2002). Bierlaire and Crittin (Bierlaire and Crittin 2004) describe an efficient algorithm for solving the problem of estimating OD. Van der Zijpp (Van der Zijpp 1997) combines gauging data with information obtained from trajectories captured using ANPR for estimating OD flow. Basically, an equation is estimated to count paths and then scattering probabilities are calculated from counting the number of vehicles on the links and paths counting.

- **Antoniou** (Antoniou, 2004) presents a methodology for incorporating information AVI in OD estimation and prediction framework; which was extended to allow consideration of any available surveillance data and then applying state models for calibration on-line relationship between speed and traffic density (Zhou and Mahmassani 2006). It could find three approaches to solution: Extended Kalman Filter (EKF) and Unscented EKF Iterated Kalman Filter (UKF) (Antoniou, Ben-Akiva and Koutsopoulos 2005) have been applied on motorways in Europe and the US. Other researchers (Antoniou 2004) develop a joint formulation of the calibration equations as transitions and actions on a state space model. Perturbation Stochastic Approximation and (SPSA), integrating techniques Advanced Vehicle Identification (AVI) and counting. Wang and Papageorgiou (Wang, 2005) present an estimate real-time traffic conditions, based on a stochastic macroscopic model in the state space, which solves using Extended Kalman Filter, this author considers dynamic monitoring and adjustment temporal parameters of the model, including as state variables to be estimated. Bañon provides a good introduction to traffic engineering (Bañon and Bevía 2000).
- **Individual Pattern Analysis and prediction.** Predicting mobility patterns cannot be a random distribution of trips within the geographic area of a city, it has to be bound to the distribution of population, of activities, and of the economic and social characteristics at particular locations. The question is that how this demand is expressed in behavioral choices today does not predict how this demand will be expressed in the future. People's current choices or preferences are flexible if the offerings of a mobility system change. Other selection is whether people, with novel choices and perhaps more flexibility in travelling at hand, will also adapt their activities, or activity locations, in order to satisfy their needs. Different techniques have been explored: Markov Chains (Gambs and Killijian 2012; Francois 2007), Hidden Markov Models (HMM) (Prasad and Agrawal 2010; Su et al. 2016), advanced clustering (Clustering based Sequential Pattern Mining (CSPM)/Sequential Pattern Mining Based Clustering (SPMC)) (Duong and Tran 2015; Avasthi and Dwivedi 2013), statistical (Park, Hong and Cho 2006) among many others from machine learning and soft computing domain.

3.11 Processing architectures

Emerging new technologies have been deployed under the term of massively parallel-processing (MPP) to address some of the issues to be tackled in this context.

- **Batch-processing/analysis.** Programming models are needed that are able to parallelize and manage the distribution of large volumes of data. Map Reduce programming model is a programming model for processing large data sets, and the name of an implementation of the model by Google. MapReduce (Hadoop, http://hadoop.apache.org/) is typically used to do distributed computing on clusters of computers. The models is inspired on the map-reduce functions commonly used in functional programming.
- **Real-Time Analytics.** Here, we can find solutions able to deal with timing constraints of data analysis process and storage. Process must be done in real-time. Two main options: CEP (Complex Event Processing) and IMDG (In-Memory-Data Grids). CEPs capture information from messages streams, databases or applications in real-time, also defines taxonomies of interrelated events, supporting their identification and automatic actions or warnings (e.g. alarms). Several solutions are currently in the market (e.g. WSO2, Esper, Siddhi, Drools Fusion and others), with different transport protocol, data format, runtime engines and pattern/event definition richness.

In addition, in the last times, we can find a new concept IMDG (In-Memory Data Grid), able to process huge datasets in real-time, using the system main memory as storage. This is specially indicated for local and volatile data, being distributed as plugins or libraries for reference applications servers. Some examples are VMWare Gemfire, Oracle Coherence3, IBM extreme Scale, Terracota or JBoss Infiinispan.
In addition, there exists several platforms specific for real-time data analysis. Some reference samples are the following: Teradata, a parallel massive processing system, based on the concept of "shared nothing", Storm & Kafka, characterized by high fault tolerance and scalability, identifies several streaming steps: collection, transportation and process and finally, Spark & Shark, with similar capabilities but implements a functional development approach. Other options are HP Vertica, Brisk, and Parstream.

- Storage Technologies. We need repositories capable of storing a huge volume of data and with distribution, scalability and performance characteristics. NoSQL databases present a simple, lightweight mechanism for storage and retrieval of data that provides higher scalability/availability, distribution and high performance in load and query time. NoSQL databases are classified according different types (Yen 2009): storage key-value (standard, cached and/or eventually consistent), data structure server (In-memory DB), object-oriented databases, document-oriented, column-based repositories and graph-based data bases. Specifically in the context of mobility analysis we find the previously mentioned products from CISCO, Siemens, Indra, among others, and also relevant European projects as: SUPERHUB (www.superhub-project.eu/), that integrates multi-modal travel planning, journey resourcing and ticket purchasing, Co-Cities (www.co-cities.eu), addressing the feedback for final users and MoveUs (www.moveus-project.eu), that integrates just mentioned capabilities with the energy efficiency aspects: measurement, prediction and savings and several other platforms defined in R&D European projects.

3.12 Complex Event Processing

The goal of Event order to Stream Processing (ESP) is to create software components capable of processing streams of event data in identifying some pattern within those streams, using techniques such as detection of relationships between multiple events, event correlation, event hierarchies, and other complex aspects.

In this context, a Rule Engine (RE) is usually referred to as a component that applies simple rules to an incoming stream of data and automatically executes the predefined consequences. By contrast, a Complex Event Processor (CEP) would take into account not only the data just received, but also the historical data (events that happened in the past and are somehow relevant). Figure 19. shows an overview of the structure of a CEP as it is currently understood.

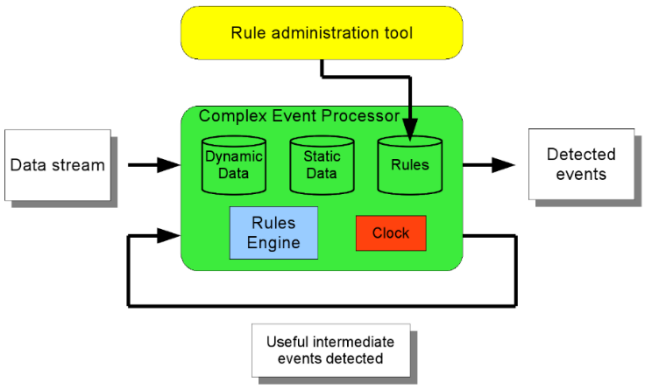

Figure 19. Interoperability.

Dynamic Data usually consists of the recent input data and possibly some useful intermediate derived data. It is stored in memory in order to apply the various rules on it. There is usually a window time frame to determine for how long the historical data must be kept.

Static Data is some other data that describes the business context and is generally unmodified during the execution. Some examples could be data describing the infrastructure of an organization, unmodifiable parameters that depend on the country's laws, etc. *Rules* are expressions, usually with an if-then structure, written in some formal language. They are permanently being checked, in order to at some point determine the existence of a new event. The *Rule Engine* is the core component. Its task is determining when the rules match their condition and acting accordingly. Optimizing the amount of computations needed to achieve this task is usually not easy.

A *Clock* is added to the diagram to stress the fact that some rules use time frames in their conditions, which means that the current time is also relevant for rule evaluation. *Useful Intermediate Events Detected* are auxiliary pieces of data that the CEP optionally stores temporarily (dynamic data) which are relevant in order to work with the rules. *Data Stream* is the input to the CEP. It often needs a pre-process in order to convert it to the CEP internal data model. Usually a timestamp is associated to it. *Detected Events* is the output of the CEP. It is usually a series of timestamped data structures that describe some kind of event that is relevant in the business context. Other components will take it as input.

Rule Administration Tool is an optional external component that establishes the rules to be used by the CEP. It usually acts before the execution begins, though sometimes it is possible to update the rules on the fly, and therefore the behavior of the CEP evolves dynamically by applying new rules to the current historical dynamic data. Some well-known examples of CEP-related software:

- **Drools** is an open-source Business Rules Management System (BRMS) written in Java that provides, among other things, some powerful complex event processing features.
- **Esper** is a component for complex event processing (CEP) and event series analysis, available for Java as Esper, and for .NET as NEsper.
- **WSO2 CEP** is an open-source component that helps identify the most meaningful events and patterns from multiple data sources, analyze their impacts, and act on them in real time.
- **SQLstream** Blaze is a stream processing package for real-time operational intelligence from the integration, analysis and visualization of high volume, high velocity machine data.
- **JBoss BRMS** is able to detect, correlate and respond to events across multiple real-time data sources with the built-in Complex Event Processing extensions to the rule language and rule engine.

- **Oracle Event Processing** (part of Oracle Fusion Middleware) is a high throughput and low latency platform for developing, administering, and managing applications that monitor real-time streaming events.

3.13 Traffic simulation for urban mobility

Simulators have been widely used in the development of intelligent transportation systems, to perform safe experiments in a controlled environment, or to predict the behaviour of new devices or new traffic strategies that involve thousands of vehicles.

Traffic simulation tools intend not only to cope with undesired events as mentioned above, but also to generate scenarios, optimize control, and predict network behaviour at the operational level. This allows a specialist to virtually modify the network topology or the traffic control strategies in order to validate the reliability of new models without any disruption to traffic in a real network.

As for traffic simulation models, there are four distinct types depending on their different granularity scales, i.e. in terms of their level of detail classification. These are the macro-, meso-, micro- and sub-microscopic modelling approaches. The macroscopic simulators are based on mathematical models describing the vehicles' flow through the network, an approach similar to the fluid dynamics.

The microscopic traffic simulation consists of a set of models representing the individual vehicle behaviour in a traffic road that should be calibrated to follow the macroscopic traffic flow patterns. Despite its complex configuration, once a good calibration is set up, the model follows the macroscopic traffic flow patterns allowing for a wider analysis on the vehicle behaviour, proving itself suitable for individual intersection optimization, e.g. traffic light planning.

The mesoscopic simulation manages to get the advantages of both macro- and microscopic simulators, combining the high level detail of entities, but describing their interactions and behaviours in a lower level, for instance in probabilistic terms. These descriptions can take different approaches, such as with vehicle grouping as a single entity with its speed calculated for each link using a speed-density based function or with individual vehicles grouped into cells that manage their behaviours being the cell responsible for determining the speed of individual vehicles (Burghout 2004). In addition to describing the time-space behaviour of the individual entities in the traffic network system, sub-microscopic simulation also known as nanoscopic simulation describes the functioning of specific parts and processes of vehicles and driving tasks, i.e. apart from a detailed description of the driving behaviour, the vehicle control behaviour (e.g. tire deformations, changing gears, inertia) is also modelled in detail in correspondence to prevailing surrounding conditions (Hoogendoorn 2001).

Microscopic traffic simulators are particularly relevant for ESTABLISH. In (Kokkinogenis 2011), a comparison taxonomy for microscopic simulators applied to future urban transport is proposed considering the following selected criteria:

- Extensibility: When using a closed-source software, its extensibility should be analysed in order to study if it suits the integration of other tools, i.e. the level of accessibility of the simulation core.
- Software License: Open-source simulators are generally inferior in features as compared to commercial ones. Nonetheless, when well documented they tend to be more flexible and rapidly extended due to community support.
- External Agent Support: The ability to use the agent technology, not only in driver behaviour modelling, but in simulation initiation, control or deployment.

- Parallelism/Distribution: To support a large traffic scenario, simulators must feature distributed processing over several cores or a computer network.
- Inter-vehicular communications (IVC): Virtual communication infrastructure support for V2V or V2I and physical restrictions simulation must be present as well.
- Interactivity: What features are controllable from simulation in run-time, and general graphical aspect:
- Level of Maturity: Whether the simulator is widely used and validated by the scientific community.

Therefore, some microscopic traffic simulators are briefly analysed in order to choose the most suitable one to meet its requirements, namely VISSIM (http://vision-traffic.ptvgroup.com/en-us/home/), PARAMICS (http://www.sias.com/2013/sp/sparamicshome.htm), AIMSUN (https://www.aimsun.com/), MITSIM (https://its.mit.edu/software/mitsimlab), SUMO (http://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883/16931_read-41000/) and MAS-T2er Lab (Ferreira 2008) .

Regarding the extensibility of these simulators, all of them offer some type of modularization. MAS-T2er Lab simulator only provides an UDP connection to control semaphoric intersections and statistical data. SUMO simulator provides an extension named TraCi, which provides statistical data and direct access to some core elements, however it is in a very embrionary level. Nonetheless, it has been extended by the TrasMAPI project to support the implementation of agents in Java (Timoteo 2010). All commercial simulators seem to fulfill this requirement, but a tougher analysis should be needed to evaluate the possibility of integrating external vehicles (e.g. from an autonomous vehicle simulator) in real-time into the simulations.

VISSIM, PARAMICS and AIMSUN are full closed-source packages, MITSIM have both the close- and the open-source variants and SUMO and MAS-T2er Lab are open-source, being SUMO the most featured and referenced open-source project with over hundred papers. The importance of simulators license is high, as in the case of modification needed on the core, only open-source ones can fully allow it.

Only MAS-T2er Lab seems to support agent-based driver behaviour simulations, with multi-connection and local information.

PARAMICS is the only off-the-shelf simulator supporting distributed computing over a network. The remaining simulators support parallel processing over all CPUs.

Inter-vehicular communications are not supported by any of the commercial variants, as they do not seem to be targeted at research purposes. Only SUMO was already modified to support this kind of simulation (Sommer 2007).

Regarding the interactivity criterion, the most 3D realistic simulators are the commercial ones, followed by MAS-T2er Lab. Only SUMO and MITSIM do not have 3D visualization. In-simulation parameter modification is only widely supported by commercial applications and SUMO. With regard to the last criterion, commercial simulators have an expected high maturity, and on the open-source side, only SUMO is being actively developed.

Table 12. summarizes the aforementioned observations.

Table 12. Features comparison of robotics simulators for agent-based autonomous vehicle simulation.

Simulator	License	Extensibility	Agent Oriented	Parallelism/Distribution	IVC	Interactivity	Maturity Level
VISSIM	Commercial	Yes	No	Yes	No	High	High
PARAMICS	Commercial	Yes	No	Yes	No	High	High
AIMSUN	Commercial	Yes	No	Yes	No	High	High
MITSIM	Both	Yes	No	Yes	No	Low	Low
SUMO	GPL	Yes	No	Yes	No	Medium	High
MAS-T2erLab	Free	Yes	Yes	Yes	No	Medium	Low

4. ESTABLISH use cases

In this deliverable, four use cases of ESTABLISH project are described:

- Optimized City and Mobility Planning (Prodevelop),
- Smart HVAC systems that ensure a healthy indoor environment (IMA),
- Rehabilitation decision support (Siveco), and
- Indoor air quality improvement at school (VTT).

The owners of these use cases are named by the company. Two more use cases will be planned and described when the national funding has been received:

- Dementia air quality (Turkgen), and
- Remote elderly lifestyle support (ISEP).

4.1 Optimized City and Mobility Planning (Prodevelop)

Description and goal

Planning of urban development and traffic greatly affects air quality and consequently living conditions of a city. Based on environmental sensor data, complemented with other data sources (e.g., building footprints, demographics, traffic information, etc.) and geo-data analysis techniques, locations and/or routes will be profiled in terms of their environmental conditions and their impact on a healthy living environment.

Implementing smart transportation is a more durable way of raising the air quality in urban areas. Three key parameters will define smart mobility in urban areas:

- Advanced traffic management system (ATMS)
- Parking management
- ITS-enabled transportation pricing systems

In the use case on Optimized City and Mobility Planning, the Spanish consortium 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 (see Figure 20.). This pilot will also enable gamification methodologies to motivate people to improve efficiency of the transport system and promote sustainable habits in the context of transport mobility.

The use case includes the development of a dashboard that displays the current status of metrics and key performance indicators (KPIs) related with it. The essential features of the dashboard product include a customizable interface and the ability to pull real-time data. The dashboard will have two perspectives, the first for the authorities and the second for the citizens.

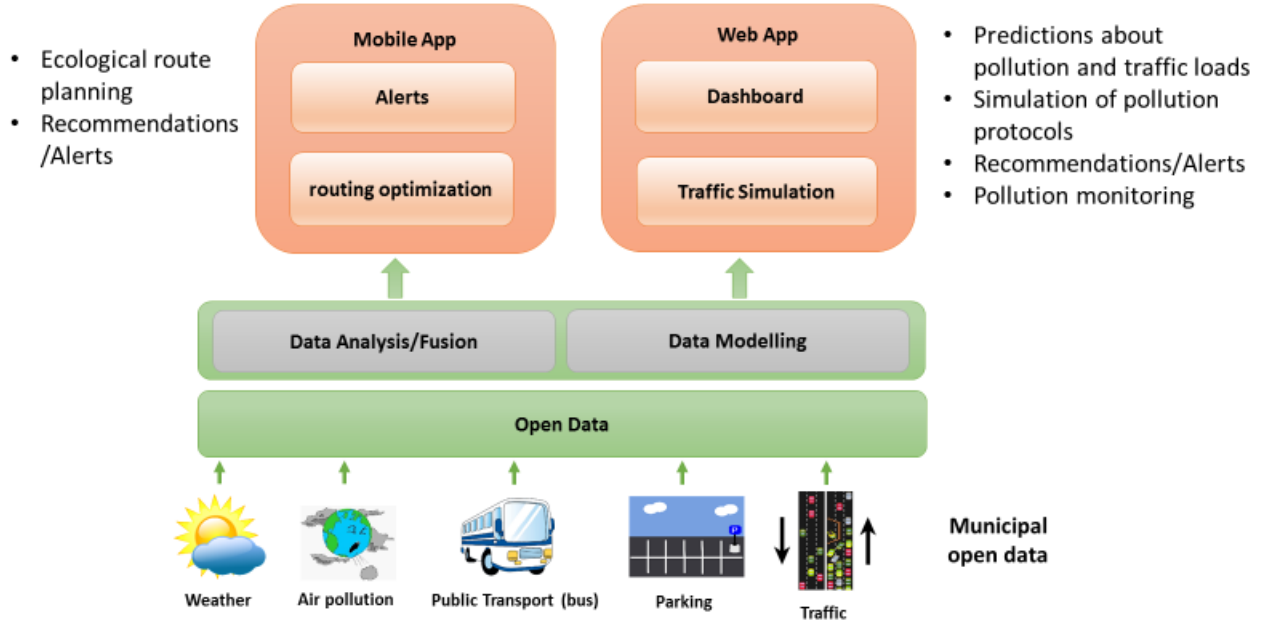

Figure 20. The idea of Optimized City and Mobility Planning use case.

The *Optimized city and mobility planning* will consider multi-modal mobility, awareness of energy usage, environmental conditions, pollution levels, travel costs and durations. The application will take into account personalized preferences and constraints, which will support individual multi-objective optimization for travelling. Multi-Modal Mobility will not only improve the quality of life of individuals and, as result, their health, but it will improve also the environment for society by reducing the air pollution produced by vehicle emissions.

Use Case Pilot

The pilot will be developed in Valencia (Spain), which is the third biggest city in Spain with 800,000 inhabitants. Valencia is investing many resources in its smart city platform and participates actively in many R&D projects and city networks (for example, Spanish Network for Smart Cities and The European Innovation Partnership on Smart Cities and Communities). Valencia is the first European city with a platform full FI-WARE compliant. Valencia smart City Platform (VLCi) is the name of its platform. Nowadays the platform contains more than 100 open data sources (http://gobiernoabierto.valencia.es/en/). These data sources are freely available for anybody, without copyright, patent or any other restriction.

Valencia city council is not a partner of the Establish project but it is committed to collaborate with the pilot from the beginning and will provide support in the use of VLCi, access to new sources that will be incorporated into its platform in the future and validation of the use case.

Current problems and limitations

The air quality in cities sometimes exceeds recommended air quality levels and can lead to negative consequences for human health. These health effects are particularly problematic in cities, because by their nature cities contain large numbers of people who can be exposed to high concentrations of pollutants. The highest pollution levels usually result from a combination of emissions near the ground and special meteorological and topographic conditions.

In 15th of February 2017, The European Commission sends final warnings to Germany, France, Spain, Italy and the United Kingdom for failing to address repeated breaches of air pollution limits for nitrogen

dioxide (NO_2). NO_2 pollution is a serious health risk. Most emissions result from road traffic (http://europa.eu/rapid/press-release_IP-17-238_en.htm).

More than 400 000 citizens die prematurely in the EU each year because of poor air quality. Millions more suffer from respiratory and cardiovascular diseases caused by air pollution. Persistently high levels of nitrogen dioxide (NO_2) caused almost 70 000 premature deaths in Europe in 2013

Road traffic is responsible for around 40% of nitrogen oxides (NOx) emissions in the EU. At ground level, the relative contribution of traffic is much higher (as emissions from high industrial stacks are diluted before reaching the ground). Of the total emitted NOx from traffic, around 80% comes from diesel-powered vehicles.

Currently, more than 220 European cities have delimited low-emission areas where non-resident access is banned or only vehicles with the lowest pollutant emissions are permitted, such as Athens, Paris, Oslo, Stockholm, Helsinki, London, Vienna, Budapest, Madrid, Lisbon, Bucharest or Dublin, among other capitals. In some cases, tolls have been established to move around the city center. As a rule, buses and trucks weighing more than 3.5 tons cannot enter.

We have already seen a significant reduction in the rate of emissions generated by a single vehicle, thanks largely to improved car engine designs, catalytic converters and improved fuel chemistry. On the horizon, we can look forward to the increasing popularity of hybrid and electric technology for personal vehicles. Most of the proposed actions are not a short-term solution and require a lot of inversion and awareness of citizens.

Since traffic is a key factor in pollution, optimizing travel within the city will contribute to the reduction of pollution. There was a time, that navigation systems and route planning focused only on providing the fastest solution to get from a source to a destination, in recent times, we are seeing all kinds of solutions that considers other factors “multimodality, Eco-routes, etc.”.

Optimizing routes, taking into account the “minimize pollution”, is a cheap and short-term solution to reduce the vehicle emissions and increase the air quality. Smart route planner should consider multi-modal mobility, user preferences, real-time information, awareness of energy usage, environmental conditions, pollution levels, travel costs and durations.

Current solutions

Navigation systems and route planners have existed for years. At first, they focused on providing the fastest route but in recent years have evolved taking into account other factors. This point will describe the evolution of the route planners over the years using as an example some of the most popular solutions.

Navigator devices (standalone GPS receivers)

Publicly available GPS devices had already been around since the early 1980s.But it was not until 2000 that precision GPS navigation became open to the public. A GPS navigation device commonly referred to simply as a GPS, is a device that is capable of receiving information from GPS[1] satellites and then to accurately calculate its geographical location and calculate the best route taking into account some configuration options. A GPS device can retrieve from the GPS system location and time information in all weather conditions, anywhere on or near the Earth. Today, most standalone GPS receivers are used in automobiles. Tomtom has been the market leader in the navigator devices since the beginning. Limitations include few configuration factors, not a multi-modal system and not a real-time system.

[1] The Global Positioning System (GPS) is a global navigation satellite system (GNSS) made up of a network of a minimum of 24, but currently 30, satellites placed into orbit by the U.S. Department of Defense

Mobile-Navigator devices (integrated in a mobile phone)

By the mid-2000's, you could add GPS functionality to an existing phone with a Bluetooth-enabled GPS receiver. However, it was not long before smartphones began to come with integrated GPS. Similar solution to "Navigator devices" but instead of using a specific device, the GPS is integrated into a mobile phone. The mobile can use the base station or cell towers to provide the device location tracking capability, especially when GPS signals are poor or unavailable. This solution has the same limitations of the previous one, but you do not need to buy a specific device, and in case of poor GPS signal, the mobile can use the cell tower to provide the device location.

Google Maps (web solution- Mobile Application)

Many phones today come with free Google Maps apps, which will provide turn-by-turn navigation between two addresses. Google Maps has a lot of features. You will get the standard turn-by-turn directions. There are also live traffic updates and it contains information about virtually every business known to man based on location. The Google Maps gives you directions for trips by car, bike, foot or public transportation. The service has local bus and train schedules for many cities including bus and stop numbers and transfer points. For longer trips, Google provides airline information including ticket price and carriers that serve your destination. It is free, available in lots of devices and you can even download maps offline for temporary use. It is possible download Google Street View to get even closer looks at destinations, businesses, and even houses. Limitations include:

- In order to obtain live (real-time) traffic update you need to be connected to internet.
- The quality/ accuracy of data is not always good.
- Minimize pollution is not considered in the route planning.

WAZE (social - Mobile App)

Waze is the second most used navigation app. It shares a few features with Google Maps. However, it is still excellent on its own. It features live traffic updates, sourced by the people who actually drive in traffic. You can make sure you are not speeding, check where police may be, and find other useful driving information. It can also find cheaper gas on your route, has location sharing with friends, and there is even a ranking system for those who really contribute to the experience. It is entirely free and worth a shot if you just want something different.

The fact that users provide the data found in this navigation app is one of the reasons behind its incredible rise in popularity among vehicle drivers. This is one reason that enabled the app to perform better than Google Maps. With this new app, drivers will get to their destinations without encountering distractions along the way. The app shows a has a pop-up that displays all types of information that other users might list concerning accidents, traffic jams and speed cams as well as the presence of police.

Waze was acquired by google in 2013, and most of the coolest features of Waze have been included in the google maps app, nevertheless the social features has not been included. Waze has the same limitations as google maps.

CityGo (Mobile App - fiware)

CityGo application can tell the user what public transport options can choose for a particular route, from where you can take the electric car (car sharing), until which bus is the best that comes to you, passing through the nearest station Public bike rental and availability. Everything is managed in real time to obtain an optimal route. The application "promotes healthy habits and respectful of the environment."

A web interface has been developed to be visualized from a control center in the City Hall, so that the authorities in Malaga can know at all times the flow of people and traffic in the city. This Information is

useful to detect needs of public services and in moments of high tourist influx and concentration of people, such as a sporting event or the development of Holy Week in the city.

This application is under development, you can download a beta version for android devices. It is being developed by the company Atos with the collaboration of the City Council of Malaga. The application uses the data provided by the open data platform of Malaga. CityGo is useful only for calculating routes in the city of Malaga.

The strength of this application is the quality of the data used to perform the calculations; these are obtained directly from the open data platform of Malaga. Limitations include:

- Only for public transportation
- Minimize pollution is not considered in the route planning.
- Limited user preferences are taken into account.

Technical development

The use case will take into consideration the following technical aspects:

- Route planner:
  - Different means of transportation
  - Mono and / or multi-modal planning.
  - Total flexibility in map definition and geographic scope.
  - Fully configurable metrics (environmental impact, time, distance, mix, etc. ...)
  - Particularization according to habits of mobility and personal preferences. Integration of public and private transport modes (buses, bicycle and car exchange), others as electric vehicle, private vehicle, bicycle).
  - Connection with environmental impact assessment and energy consumption systems
- Precise prediction and forecasting of traffic loads and pollution, simulation of pollution protocols based on based on the information provided by the Traffic Simulation Platform.
- Recommendations for mobility and alerts about pollution in real time based on the fusion of different open data (such as air pollution, weather, traffic, etc.) using big data and machine learning techniques.
- Environmental impact assessment.
  - Estimation of environmental impact “air pollution” based on traffic, topographic and meteorological information.
  - Indicators by zones.
  - Asynchronous notification of alarms against thresholds.
- Integration of different public data sources useful for the use case.
- Consideration of user preferences, fuel consumption and contamination of the user's vehicle and health problems.
- Provision of a web solution for city authorities and a mobile application for citizens.
- Develop a dashboard to analyze the information in real time.

Figure 21. presents a first approach for the implementation of "*Optimized City and Mobility Planning*" Use Case.

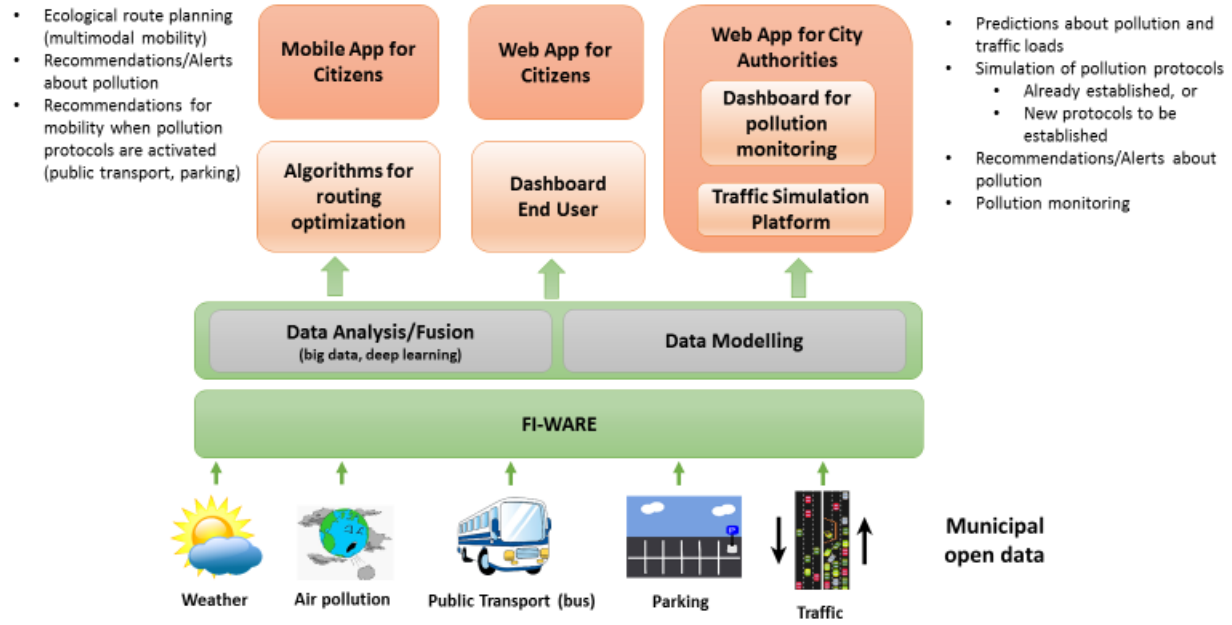

Figure 21. First approach for the implementation of "Optimized City and Mobility Planning" Use Case.

Technical specifications of the use case are presented in Appendix 1A.

Stakeholders and value realized

These are the stakeholders and value realized:

- Sensor manufacturers
- City
  - Provides a new service for their citizen
  - CO_2/NO_2 pollution decrease on a city level
  - Energy savings on a city level
  - Get feedback and metrics from the use of the application
- Open Data Provider
  - New data can be included in the “route planner”
- Public Transportation
  - Get feedback from the use of the application
- Service provider
  - New services based on the data generated from the use case
- Consumer / citizen
  - Route customized for each user
  - Better quality of life
  - Obtain pollution/restriction alerts

Impact

The implementation of the use case will provide a series of benefits that will affect the increase in the quality of life of the citizens of Valencia. Thanks to the use of the application, the users will get several benefits such as reduction in average travel time in intra-urban movements, optimal routes personalized taking into account their preferences, mobility problems or health problems. The application will promote the use of greener and healthier means of transport, reducing energy consumption and pollution in cities. In order to carry out this system, the system will use user information and data provided by different real-time data sources as well as weather forecasts. Through complex algorithms and forecasting models, the system will be able to predict future traffic congestion and elevated levels of pollution.

In addition, the system will provide feedback for cities and citizens. The application will show indicators on the use of different means of transport and roads. This information will be useful to detect needs of public services and to provide continuous improvement in these services.

The services for profiling geographic areas and routes in terms of their environmental conditions for city and route planning purposes will have the following impact:

- Travel time reduction for individuals
- Energy savings on a city level
- CO_2/NO_2 pollution decrease on a city level
- Reduction in overall queue lengths at public services
- Proportion of people using Multi-Modal Mobility services

Challenges

The Optimized city and mobility planning will consider multi-modal mobility, awareness of energy usage, environmental conditions, pollution levels, travel costs and durations. Although there are many web pages and mobile applications that calculate the route between two points, there is no solution that obtains the optimal route taking into account the user preferences and the data sources used in this project.

The application uses many user information and preferences, including fuel consumption and contamination of the user's vehicle and health information (allergies, mobility problems, etc.). The system uses different type of information, updated in real time to calculate the routes: traffic, weather, pollution, public transportation schedule and routes, parking places, public bike, pollen, etc. These data sources come from the VCLi platform.

The main challenges of the system are:

- Develop an algorithm that calculates the optimal route taking into account all the information.
- Precise prediction and forecasting of traffic loads and pollution and simulation of pollution protocols
- Recommendations for mobility and alerts about pollution in real time.
- Integration of different public data sources useful for the use case.
- Real time dashboards that manages a big quantity of information.

Partners

The possible partners for this kind of solution include:

- Sensor providers
- Open data providers
- Mobile application developers

- Software company / Research institution with the following expertise (routes calculation algorithms, Geographical information system, Prediction Models)

4.2 Smart HVAC systems that ensure a healthy indoor environment (IMA)

Description and goal

HVAC is of interest for many years and most of innovative technological trends in microelectronics and communications are addressing such systems. More over the health of population that is on top of EU initiatives is also strongly affected by HVAC systems specifically when we speak about COPD diseases.

For example the last seminar of REHVA (Federation of European Heating, Ventilation and Air Conditioning Associations) taken place on the April the 4th in London (http://www.rehva.eu/publications-and-resources/event-presentations/rehva-annual-meeting-2017-seminar-04042017.html) strengthened several HVAC challenges, value chain and recant HVAC orientation towards energy efficiency, new HVAC components and considering stakeholders who can benefit from innovative HVAC systems.

One of the current strategy of the European Commission is Energy Efficiency where Heating and cooling is one of severe initiatives since " 70.6% of energy consumption was used for space and industrial process heating, 26.7% for lighting and electrical processes such as machine motors, and 2.7% for cooling. In EU households, heating and hot water alone account for 79% of total final energy use…." (https://ec.europa.eu/energy/en/topics/energy-efficiency/heating-and-cooling).

In the industrial and research fields, many people have developed the IoT such as Indoor Air Quality(IAQ) and Outdoor Air Quality devices installed in indoors and outdoors to measure the environmental information related to the air quality. Then the cloud server can store, manage, and analyze the environmental data collected by IoT devices so that it will extract the enhancement indexes to improve the operating ability of the air cleaners.

The Czech pilot will develop such a HVAC system that will also 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.

The Korean pilot will contribute to developing a healthy indoor environment in the buildings that can automatically adjust its conditions to better temperature, air quality, humidity, etc. using the results analyzed by the environmental data. It will achieve the solution to generate the well-being, safe, comfortable indoor spaces for the people, especially the senior and the children.

Current problems and limitations

Building of new energy efficient homes initiated discussions about quality of living in such homes in terms of the air quality. New technologies used in process of construction usually restricts natural air circulation and air exchange. Tight windows and good insulation help to decrease costs although inhabitants have to be aware of need of the air quality checking. While the energy savings are of great benefit to everyone including the environment, there is a trade-off if homeowners are not educated about it. Homeowners have to be aware of:

- The moist, because tight home has absolutely no way to get rid off.
- Stale air and even toxic air that are built inside the home even from ordinary activities of everyday living.
- Allergens and odors building up.

All those aspects cause the raising of irritation and illness.

General recommendations for passive improving of air quality in home are:

- dust and vacuum often
- use of eco-friendly products for cleaning
- run a humidifier
- don't smoke
- have plants in rooms.

Many people suffer from find dust at warning level that can adversely affect the human health due to the environmental pollutions. Especially, indoor air quality is polluted by major household activities such as cleaning, cooking, etc. and outdoor air pollution. It is the main reason that the air quality driving equipment such as the air purifiers has been widely used in homes and offices. The air condition system, which is widely used for improving the air quality in indoors, is operated only by the index of the air pollution around the air purifier. It is hard to achieve the air quality improvement for the entire indoor space. Therefore, the user requires the following techniques for the reliable improvement of indoor air quality.

- Accurate air quality measurement for indoor spaces
- Improvement of air quality according to living pattern of indoors residents
- Determine whether to ventilate according to outdoor air pollution degree
- Replacement of the filters of air purifiers by measuring its pollution degree

Current solutions

Following Figure 22. shows, how air quality monitoring should be integrated at home environment management, in order to ensure elimination of negative aspects of tight homes and to contribute to quality of living.

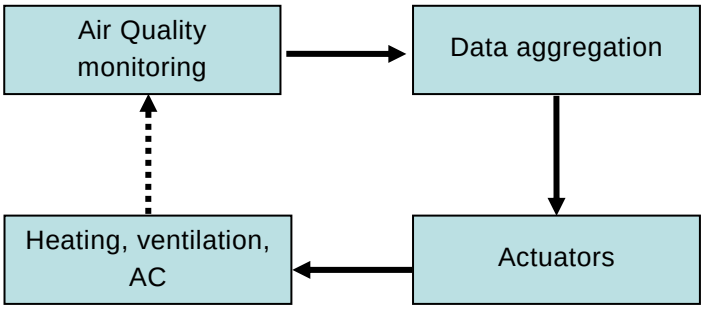

Figure 22. Usual management of home/office environment.

Air quality monitoring systems and sensors

Present trends are to use off-the-shelf sensor modules supplied by specialized producers. Modules differ by the measured parameters; they can be built as single chip modules or more complex modules. For example, sensors of the temperature, humidity, pressure are usually one-chip-built component in one case. More complex sensors, like sensors for gasses, are composed of more chips, optical elements and they are built in a proper case – so formed separate module for using in the final product.

Specialized producers select proper sensor, measuring method and control logic block. Measured magnitude is accessible at output interface, usually in digital value. It ensures the stability and accuracy of measurements. Final producer of measuring device, using sensor of specialized supplier, cannot negatively affect the quality of measurement.

Temperature

Recent temperature sensors are usually composed of following building elements:

- Sensor
- A/D converter
- Control logic
- Serial interface (most often I2C, SPI, SMBus).

Range of measuring would be from -40 to +100°C, accuracy +/- 0,3°C, resolution 0,04°C at 12 bits A/D converter.

Relative humidity

Relative humidity sensor device scheme is similar to temperature sensor, only the difference is the sensing component as a data input. Also, the interface is similar.

Range of measuring would be from 0 to 100%, accuracy +/- 0,3°C, resolution 0,03% at 12 bits A/D converter.

Air pressure

Relative humidity sensor device scheme is similar to temperature sensor, only the difference is the sensing component as a data input. Also, the interface is similar.

Range of measuring would be from 50 kPa to 115 kPa, accuracy +/+1 kPa.

CO_2

Table 13. shows the possible health effects of CO_2 concentration. [Ref.: Serinus (tm) CO_2 Gas Sensors PYM 122, Air Quality, Air Safety]

Table 13. Levels of CO_2 in air and possible effects.

CO2 level	Physical Effect to Human Being	
380 ppm	0.038%	**Natural environmental concentration**
1000 ppm		**Pettenkofer number (comfort limit)**
1500 ppm		**Typical level in unventilated rooms** **Maximum work-place concentration (MAK) → no health impact under**
5000 ppm	0,50%	**Continuous exposure** **Short term (15 minutes) maximum exposure value- heavy breathing**
20000 ppm	2%	**Recovery after 5 mins of fresh air exposure**
40000 ppm	4%	**Exhalation concentration - CO_2 narcosis**
50000 ppm	5%	**Headache, giddiness, unconsciousness, hyperventilation**
80000 ppm	**8%**	**Lethal concentration within 60 minutes**

Current system solutions developed in IMA and taken as background of ESTABLISH platform is depicted below (see Figure 23.) and explains the basic architecture of simple local air quality monitoring system. Two main system components are sensor devices and data aggregator. The data aggregator is

a device, which forms a gateway between a ZigBee Pro data collecting network and LAN. Data from sensors are handled by corresponding processes which have an access to a local data cache as well as to a remote database (managed by BUT). Furthermore, there is also a possibility to provide a web services to present data from local data cache on request.

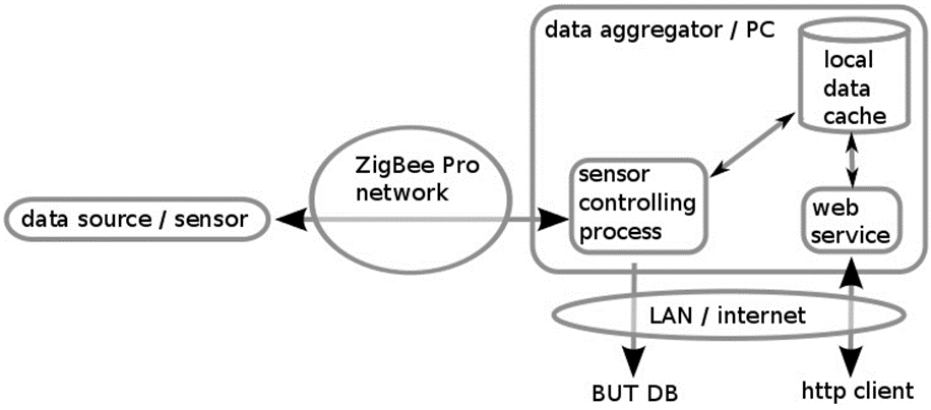

Figure 23. Current system solution in this use case.

The Web-based interface for MAS DB data management has been developed on a Microsoft Silverlight platform. Applications developed on this platform offer a visually attractive and intuitive user-friendly interface to many database systems and fulfil all requirements to the user control comfort and data visualization. The Web-based interface for MAS DB data management represents a GUI (Graphical User Interface) which runs on the core application described in Deliverable 3.7. The background core application processes all the data and provides the requested functionality to the GUI as seen below (Figure 24.), were user records and its chart are displayed.

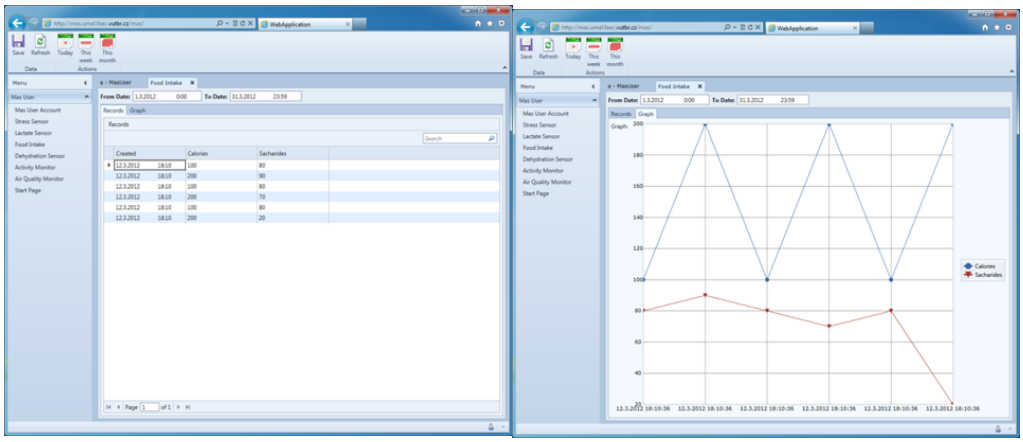

Figure 24. Visualization of user records and its chart.

Indoor Air Quality (IAQ) and Outdoor Air Quality (OAQ) devices

IAQ and OAQ devices integrate many kinds of sensors to measure the various environmental information into one hardware module. In addition, they can include the wireless communication module to transmit the measured information.

Management system of IAQ/OAQ devices

The cloud server platform can extract the key information by analyzing the environmental information in the form of the big data. Moreover, it can improve the accuracy and the usability of IAQ/OAQ devices by operating with the structural map of the living space.

Analysis of the environmental data

Analysis of environmental information that finds out the indicators to improve the air quality by classifying the user life patterns and detecting the air pollution sources. In addition, it can improve the effectiveness and minimize the management cost of air cleaners by notifying the replacement cycle of filter parts, which are the core of the air purifier, through analyzing the contamination degree of the environment.

Interconnection with the air purifier

The design of the protocol that is used for the program to interwork the intelligent management system of the air quality and the air cleaner. Moreover, the field test can improve the reliability of the developed system for the commercialization in the practical applications.

ETRI and Coway suggest the intelligent air quality management system as the current solution to improve the air quality based on the ESTABLISH framework as shown in Figure 25. The IAQ/OAQ devices can measure the environmental data affecting the air quality for the work package 4 in ESTABLISH and deliver the data to the server platform through the WiFi network. The server platform collects and manages the environmental data to monitor the air quality and to extract the pleasant index of air quality for achieving the work package 5 in ESTABLISH. The analysis result of environmental data used for improving the air quality by interconnecting with the air purifiers for the work package 6 in ESTABLISH. Finally, ETRI and Coway promote the commercialization of the developed system that can provide the comfortable air condition to users through integrating the IAQ/OAQ devices, the server, the analysis algorithm and the air purifiers.

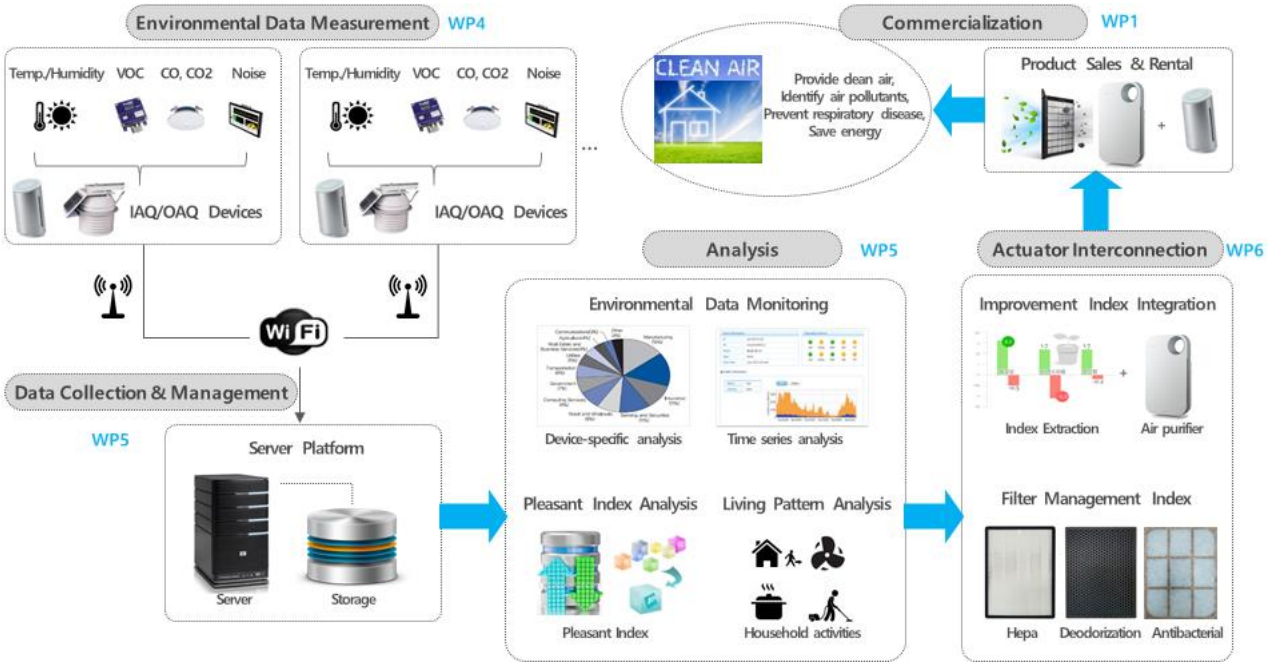

Figure 25. System overview of intelligent air quality management system based on IAQ/OAQ devices.

Technical development

Technical development will be concentrated to both new use case components and innovation of currently available components. Depend on partners' objectives the development is directed towards

- sensors, their interfaces and data gathering (gateway) (IMA),

- data processing on the cloud framework, algorithms, assessment, data interpretation (CU),
- control and act of HVAC components based on sensor data assessment (DEKPROJEKT).
- development of IAQ / OAQ devices, interconnection with the air purifier (Coway)
- management system of IAQ/OAQ devices, Analysis of the environmental data (ETRI)

The use case will demonstrate components running in the whole value chain when special interest is given on feedback control functions and services. Even though fully automatic process is intended, all the stakeholders will be able to influence the component behavior via their mobile devices and user profile can be employed. Technical specifications of the use case are presented in Appendix 2A.

Stakeholders and value realized

Use case stakeholder were already identified as follows, the order reflects the significance of the use case for them:

- residents either in working or home environment
- appliances and building components designers and producers
- health care data processing companies
- building energy manager and energy producers.

Impact

Providing superior indoor air quality (IAQ) can improve health, work performance and school performance, as well as reduce health care costs, and consequently become a source of substantial economic benefits. While some see energy efficiency and IAQ as contradictory goals, an integrated design suggested in the Use Case can bring on the market an affordable high performing control system for buildings or flats that will bring savings in energy consumption (up to 15%) while maintaining good IAQ. All stakeholders yearn for healthy environment either if they stay there as residents or if they design new components of HVAC. Building up a new subsystem - environmental sensor network - on top of ID system platform will extremely speed up introduction of new air quality services and applications within the current systems since the main system components and communication channels are already verified on presently running ID and building services. Facilitation of the new sensor subsystem implementation makes faster market access possible and brings essential competitive advantage.

Moreover, ESTABLISH can contribute to developing the technologies and building the business markets for becoming the better fresh and safe places in the home and buildings, through connecting the information between the human and the IAQ/OAQ devices. As the use cases especially in Korea, ESTABLISH will assist the people residing in the indoor environment for a long time (especially, children, patients, workers, senior citizens).

Challenges

Use case partners perceive the challenge at continuous development of core competences on system integration, new global services and cloud solutions deployment within IoT architectures. On top of innovations supported by European projects achievements, use case participants build up new business activities at growing market segments linked to important customers. The outcomes of ESTABLISH project will address new HVAC services development for home or working environment and exploit achievement of previous project. In ESTABLISH project IMA and DEKPROJECT will integrate scientific results (predictions, algorithms, self-learning processes) in their product. In addition, ETRI and Coway will share the technical experiences and the business networks in order to implement the intelligent air quality management system and succeed its commercialization through the ESTABLISH project.

Partners

Partners involved in the use case are IMA, CU, DEKPROJEKT (all CZ), ETRI and Coway (all Korean).

4.3 Rehabilitation decision support (SIVECO)

Description and goal

The ESTABLISH project appeared in the context in which wireless technologies have opened up new possibilities and applications in increasingly diverse areas of the medical field. Using medical devices together with communication technologies to monitor certain conditions and/or symptoms has laid the foundation for the integration of M2M technology in the medical industry, a phenomenon referred to as telehealth and e-health. eHealth services and mHealth applications contribute to improving the efficiency and coverage of programs for personal health monitoring, also facilitating communication between patients and medical professionals.

The Romanian *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.

The Romanian ESTABLISH project's goals are:

- to monitor health parameters to constantly improve the health of the population through rehabilitation and spa care, specifically targeting the patient's functional aspect of integration in everyday life, environment and work.
- to develop a decision support system and services based on the outdoors environment parameters and indoor location.
- to reduce operations costs and improve quality of the services provided.

We will use a first group of students from high school and a second pilot group of adults. We will register their physical activities, the cardiac rhythm, the burned calories, in order to find links between them and the air conditions.

Current problems and limitations

The study of the resources shows that specialists in medicine and ecology have established a direct link between environmental degradation and the increasing number of people suffering from allergies, asthma, cancer and other diseases. The main pollutants that act negatively on the human body are: nitrogen oxides, sulfur dioxide, ground-level ozone, carbon monoxide, formaldehyde, phenols, particulate matter (PM10 and PM2.5).

Current solutions

The proposed solution will have the following objectives:

- monitoring the risk factors, determined by analyzing the information collected from the sensor network and the vital parameters of the patients in order to help improve their health with the help of recovery/keeping fit methods
- the optimization of the recovery methods that focus on the functional aspects of the integration of the individual in society and in the workplace.

The solution is based on a flexible architecture with a scalable structure, which can serve the heterogeneous component of different types of sensors, protocols used in IoT and cloud computing technologies, technologies with sufficient security and privacy elements.

The system developed in this project will have two major components: the sensor network and software applications (web based and mobile). The sensor network will allow the monitoring of the information retrieved by sensors of various types both from the environment (indoor and outdoor) and as well as the monitoring of the subjects' vital information (wearable sensors). This information will be transmitted to the software solution (web and mobile applications) that will allow the analysis of data in order to establish the optimal conditions for effort and the required level. The network of sensors will collect data specific to each type of sensor:

- environmental sensors collect information such as temperature, pressure, humidity, the concentration of the various components of the air that may affect the quality of life (nitrogen monoxide, carbon dioxide, ozone, etc.)
- body sensors will collect information such as: blood pressure, heart rate, blood oxygen level, temperature, motion information related to the subject, the effort, the type of movement, the accuracy of the movement and its amplitude. The Fitbit bracelets will also be included here.

This information will be transmitted wirelessly to software solutions for easy access and complex analysis. Applications will provide relevant information so that the user can act upon environmental conditions.

Thus, software solutions, both the web and the mobile ones, will have a friendly interface and will be easy to use so as the end users (doctors, therapists or clients/patients) to have easy access to the data stored in the application.

Technical development

Technical specifications of the use case are presented in Appendix 3A.

Stakeholders and value realized

Exposure to high levels of air pollution, but also the climatic conditions in interior spaces have a severe impact on the health and welfare of each of us. These factors lead to an increased intake of drugs and a growing number of hospitalizations, especially for vulnerable groups such as the elderly, patients with respiratory or cardiovascular problems, children, and to a small extent, employees working in offices. Although the policies implemented in recent years have brought improvements in air quality, pollution is still the main danger to people's health. As an environmental factor, its consequences are high costs in health care, unhealthy working environments and an alarming number of premature deaths in Europe.

The Romanian pilot aims to offer people an innovative system, a source of information on the environmental conditions in which they conduct their daily activities and by analyzing this information and the decision support module to ensure the independence of people in vulnerable groups (children, women, elderly) from health care professionals specialized in recovery and rehabilitation.

The stakeholders are: Premium Wellness Institute and National Institute of Rehabilitation, Physical Medicine and Balneoclimatology Bucharest, schools, parents, companies, etc.

Impact

The project aims to offer people an innovative system, which will be their source of information on the environmental conditions in which they conduct daily activities. By analyzing this information and the decision support module, it will ensure the independence of vulnerable groups (children, women, elderly) from the health care professionals specialized in recovery and rehabilitation.

This system will reduce the negative effects that environmental conditions may have on the health of the people in these types of groups and will increase their independence and wellbeing.

Air quality monitoring studies conducted in Romania draw attention to the risk areas and hazardous substances, highlighting the link between air pollution and the effects of this phenomenon on health.

The number of those that do outdoor physical activities (jogging) or in special gyms is on the rise. Apart from those who intend maintaining their fitness levels and preventing obesity, there are people who have been prescribed physical activities related to recovery / the restoring of motor functionalities, both outdoors and indoors.

Challenges

The main innovative and technological results of the ESTABLISH project will be:

- developing reliable methods of self-knowledge and self-adaptation tuned to the specific field data collection sensor networks based on narrow bandwidth and limited use of energy consuming devices.
- using prescription analysis - ESTABLISH demo applications will collect extensive data from various sources, including recorded data.
- designing data formats and protocols that support the management of the information flow coming from various heterogeneous sensors, thus enabling the creation of a platform for the exchange of relevant data models will be designed so that they can consider sensors for modelling business processes.

Partners

BEIA, Premium Wellness Institute and National Institute of Rehabilitation, Physical Medicine and Balneoclimatology Bucharest.

4.4 Indoor air quality improvement at school (VTT)

Description and goal

The *Indoor air quality improvement at school* pilot will study the use of a variety of indoor sensors and wearables combined with users' personal feedback and environmental sensing information to provide a healthier living environment for pupils, teachers, and other staff members, see Figure 26. The idea is to learn from sensor data by means of combining different data streams and applying environmental models, machine learning, data mining and supported by big data ICT. This will enable optimization of power provision and self-adaptive HVAC control and air purification in the house. The users can get personalized recommendations based on measured data and feedback got from the users on the impact of air quality.

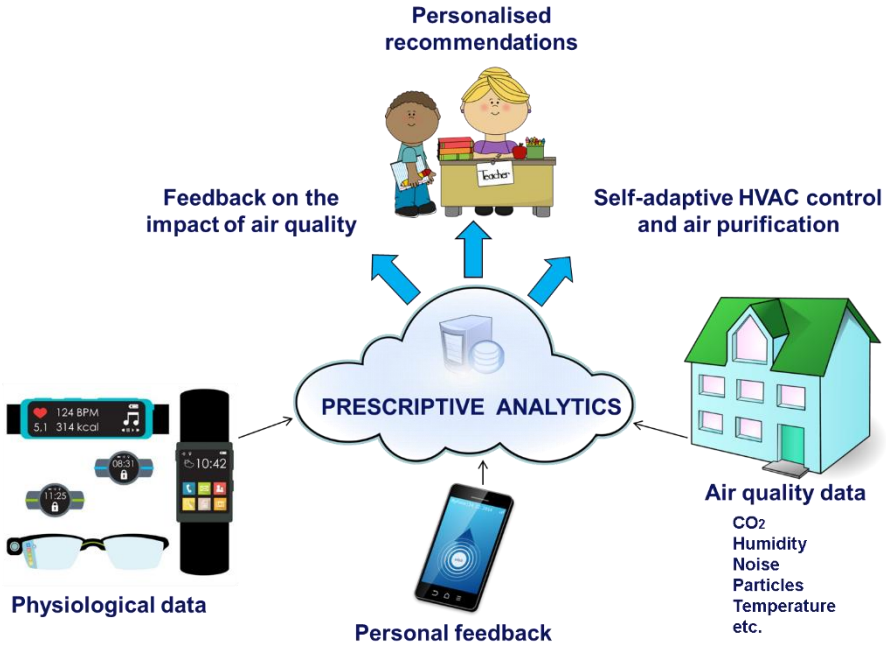

Figure 26. Idea of the Indoor air quality improvement at school.

Pilot. The pilot will be performed at a school in Finland. Certain rooms at school will be equipped with air quality control systems, determined in technical specifications, based on the state-of-the-art report in chapter 2.1. Sensors for indoor air quality measurements will include: temperature, noise, humidity, air pressure, carbon dioxide (CO_2), nitrogen dioxide (NO_2) and TVOC (total volatile organic compounds).

In the pilot, the staff members of the school will receive suitable sensors for monitoring their health condition (physiological data) like heart rate (HR) or heart rate variation (HRV), activity and sleep. They are able to provide feedback about their health, their feelings related to the air quality inside the house, by a mobile application and to control the system. The feedback data and data from the wearable devices and sensors will be collected to Azure cloud platform; stored, pre-processed and analyzed based on the time stamps. Based on the real-time analyses, users could get e.g. visualizations about the air quality (data from the sensors) compared to their health via the application. The need and possibility to utilize the air purifiers in the pilot at school will be evaluated.

The pilot consists of two phases. During the first phase, data will be collected to establish the connection between perceived and measured health and air quality. The second phase will utilize the understanding gained from the data of the first phase to provide the users with insight on the dependency of their health on air quality and allow for smarter control of their environment. The pilot will be implemented at the school in autumn 2018.

Test environment. The setup of the pilot will be tested beforehand in VTT's premises in Oulu in autumn 2017. Ten offices (ten testee) and meeting rooms will be equipped with indoor air quality control systems (same sensors than mentioned above) and testee will get the wearable sensors to monitor their physiological data. The application will be developed during the test period so that in the first phase we will only collect the feedback about the air quality from the users, as "How do you feel". Other elements will be included before the real pilot after enough data has been gathered. Available TePi environment will be utilized in the pilot including Azure cloud platform to data handling, as it has been planned in the real pilot. An unique TePi piloting environment at VTT comprises of an Industrial Internet (IoT) platform, a data management & analysis platform and a demonstration facility. The Industrial internet platform exploits the latest advancements of 5G, IoT and Cyber Security technologies and methods. It consists of

electronics pilot production and smart building research environments, and introduces 5G infrastructure to support the digital service and application piloting. In addition, related data management and analysis platform includes a server cluster for computationally intensive tasks for enabling lightning fast parallel computation, *e.g.* training deep learning neural networks and fast processing of huge datasets.

This setup will be tested during several months to collect data and to be sure, that it will work properly in the pilot at school. Feedback from the testee will be utilized iteratively during the pilot to make some modifications before the real pilot at school if needed.

User involvement. To understand the needs and expectations of the users, the work will be started to involve the users (VTT workers and staff members of the school) to the innovation process. Users will be interviewed to understand the present state and how would they like to improve the air quality and service related to it. These results will be utilized when planning the application and future services for the users.

Current problems and limitations

During the past decades, various air purification devices have been developed to resolve indoor air problems. These devices can be used to eliminate pollution, to a degree. However, often the root cause of unhealthy indoor air is not detected. Low quality indoor air causes for example tiredness, headache and stuffiness. Extended periods of polluted air cause diseases and shorter life span (WHO, 2015). The Clean Air for Europe project estimated in 2000 that there was 369 000 premature deaths in Europe and 1 300 in Finland because of air pollution and the number is expected to increase (Watkins et al., 2005).

Typically, air quality measurements need to be ordered from suppliers and it takes several months them to deliver results. Thus, inhabitants do not get any direct feedback about the indoor (or limited outdoor) air quality and the results might be inconclusive.

An R&D challenge is to convert near real-time environmental sensor data into actionable information for users to provide a healthier and safer environment. Current application barriers include the availability of relevant sensor data, data platforms and analysis to extract useful information. This project aims to provide the means to utilize environment and wearable sensor data for a better understanding of air quality, and enable the applications that can ultimately improve everyone's life.

Current solutions

IoT sensors are widely being installed in our environments and are able to provide the new methods for acquiring information about the ambient conditions, such as climatic conditions, pollution levels and infection sources, and the occurrence of emergency situations. While traditionally environmental sensors are already being used for specified tasks, *e.g.* for controlling household heating in a simple static way, their full potential lies in collecting information on a sensor network level (including personal wearable sensors) in such a way that it allows control in an adaptive and non-predefined manner. Current implementation barriers include that the application of wearable sensors that may provide more useful information on a personal level is rare, platforms for sharing data (e.g. temperature, relative humidity, CO_2 / CO / NO_2 concentrations) among environmental sensor owners are limited and no useful interpretation of the collected data are available. So, what these data signify or tell us to do remains unclear. This project aims to bridge this gap between such data and their meaning.

There is a growing movement towards person-cantered services, which combine personalization, choice and user engagement. By providing vulnerable groups with information on their environmental conditions and giving them tailored advice on their options, they can monitor and enhance their health more independently. Thereby, health effects as a result of environmental conditions can be reduced and personal independence and wellbeing can be enhanced.

Technical development

In the *Indoor air quality improvement at school* pilot a variety of indoor sensors (temperature, noise, humidity, air pressure, carbon dioxide (CO_2), nitrogen dioxide (NO_2) and TVOC) will be implemented at school environment. The data from indoor conditions monitoring will be combined with outdoor weather and air quality sensors (temperature, air pressure, humidity, CO2, precipitation, wind speed and direction, particulates, NO2, SO2, CO, O3) and wearable sensors (heart rate (HR) or heart rate variation (HRV), activity and sleep).

The feedback data and data from the wearable devices and sensors will be collected to Azure cloud platform; stored, pre-processed and analyzed based on the time stamps. Based on the real-time analyses, users could get e.g. visualizations about the air quality (data from the sensors) compared to their health via the application. A mobile application for the users to visualize the results and control the indoor air quality will be created based on their needs. The need and possibility to utilize the air purifiers in the pilot at school will be evaluated.

More detailed technical specifications of the use case are presented in Appendix 4A.

Stakeholders and value realized

Stakeholders that may gain new business with the Establish solutions and that are needed for developing the pilot include sensor and other device manufacturers and data service providers. Sensor manufacturers gain better insight of potential of sensors in increasing wellbeing and new customers. Device manufacturers can make their devices betters (e.g. air purifiers) utilizing the monitored data. Data service providers (integration, analyzing, storage, privacy…) have an opportunity to bring new and existing services to this growing application area.

Facility service providers have new business possibilities to offer comprehensive service for the users of the property. In general, better view to the indoor air helps the work of facility services in making correct decisions on time. For the owner of the property Establish solutions offer a direct communication channel to the users of properties and better understanding of how to create healthy environments. Consumer / user of the property gain accurate information about indoor air quality and the possibility to affect the indoor circumstances.

Impact

The goal of this pilot is to research, co-innovate and develop a cloud-based air quality measurement, analysis and feedback system that provides information on the impact of indoor or limited outdoor (*e.g.* mine, tunnel) air quality on people's health and wellbeing. The main application scope of ESTABLISH is products, services and solutions on an individual level utilizing environmental sensors (air quality, temperature) and combining the environmental input with other data sources, for example personal wearable sensors. This pilot will enable to go from *monitoring* the indoor or limited outdoor to *managing* the environmental conditions on a personal level and thus improving the quality of life, reducing health costs and supporting vulnerable groups such as children.

The technological developments will lead to a wide range of new services and products that are all based on environmental sensors. Thus, ESTABLISH will create business opportunities for sensor manufacturers, service providers (*e.g.* facility management companies), software developers, health organizations, health insurance companies and HVAC manufacturers.

Challenges

One of the most difficult challenges of the pilot is perceived to be data privacy. Consumers may not be willing to share their personal information or feedback. Furthermore, the rate at which information is

gathered from consumers may be critical. The size and type of the health sensors is also important. In order to benefit from the health monitors the consumers should be willing to wear those long periods of time.

Technically, reliability and integrability of data is critical is order for it to be useful in analysis. Data will be integrated from multiple, heterogeneous sources. Furthermore, the system should be able to distinguish unreliable data from reliable. Sensors may become unavailable or broken. Data should be also protected from unauthorized access.

Partners

The potential partners for this kind of solution include sensor manufacturers, device manufacturers, database analytics companies, data security companies, and mobile application developers.

5. Use case requirements

5.1 Optimized City and 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. This pilot will also enable gamification methodologies to motivate people to improve efficiency of the transport system and promote sustainable habits in the context of transport mobility.

The pilot will provide a dashboard to display metrics and KPIs. The essential features of the dashboard product include a customizable interface and the ability to pull real-time data.

After the meeting between the Spanish consortium and the city council of Valencia, the scope of the pilot and the requirements have been defined.

Use case requirements are presented in Appendix 1B.

5.2 Smart HVAC systems that ensure a healthy indoor environment

The use case “Smart HVAC systems that ensure a healthy indoor environment” will consist of HW and SW components, customer services and applications, running on specific cloud platform. The main requirements on HW components are identified, partners will develop various sensors monitoring indoor air-quality, typically temperature, relative humidity and CO2 (IMA); Dekprojekt will develop ventilation actuators. The HW components will be provided by control SW. As for the SW components; CU will develop edge cloud server and algorithms processing big data from sensors resulting in control processes at ventilation systems considering customer profiles and desired optimization schemes.

The use case will provide human interface and customer application to allow customer either feedback monitoring or/and active impact on the algorithm when optimization processed. Dekprojekt acts within the use case in role of system integrator and also in role of technology user since its company business model.

All partners will participate at requirements elicitation and testing/validation. A test components of the use case will be deployed at IMA and Dekprojekt premises.

Use case requirements are presented in Appendix 2B.

5.3 Rehabilitation decision support

The pilot study which developed in Romania will use data from ambient sensors (air quality, atmospheric pressure, temperature, humidity) and the biometric, physiological and behavioral sensors, (heart rate and respiratory number of steps) to provide patients in clinical rehabilitation/ keeping fit programs with decision support tools related to behavior and treatment options.

The project will test mobile and web applications that offer users/clients data regarding decision support and location information. The system will be integrated with handheld devices like Fitbit bracelet that record the steps, distance and calories burned during the day and monitor sleep quality at night.

The first project activities within ESTABLISH started in November 2016 with the study of sensors and applications that will monitor the physical activity, physiological parameters and environmental conditions in which the running / walking exercises meant to keep fit the students of "Grigore Moisil " National College in Bucharest are carried out. These extracurricular activities are part of the thematic

area of interdisciplinary STEM designed to revive the interest of Romanian students for the combined study of science, technology, engineering and mathematics.

Use case requirements are presented in Appendix 3B.

5.4 Indoor air quality improvement at school

In the *Indoor air quality improvement at school* pilot, data from of indoor and outdoor sensors and wearables is combined with users' personal feedback and environmental sensing information to provide a healthier living environment for pupils, teachers, and other staff members. The idea is to learn from sensor data by means of combining different data streams and applying environmental models, machine learning, data mining and supported by big data ICT. This will enable optimization of power provision and self-adaptive HVAC control and air purification in the house.

The users have a remarkable role during the pilot. First, they are informed about the idea of the pilot and their role in it. Users are involved in the planning of the pilot and the future mobile service to really understand their needs and expectations. The idea is that the users can get personalized recommendations based on measured data and feedback got from the users on the impact of air quality. Feedback is collected regularly during the pilot by using different methods. More detailed use case requirements are presented in Appendix 4B.

6. Conclusions and next steps

Four different use cases have been defined for ESTABLISH project: 1) Optimized City and Mobility Planning, 2) Smart HVAC systems that ensure a healthy indoor environment, 3) Rehabilitation decision support, and 4) Indoor air quality improvement at school. Two more use cases will be planned and described when the national funding has been received. The first use case, Optimized City and Mobility Planning, concentrates on outdoor air quality; planning of urban development and traffic greatly affects air quality and consequently living conditions of a city. Based on environmental sensor data, complemented with other data sources (e.g., building footprints, demographics, traffic information, etc.) and geo-data analysis techniques, locations and/or routes will be profiled in terms of their environmental conditions and their impact on a healthy living environment.

The other three use cases concentrate on indoor air quality. The Smart HVAC systems that ensure a healthy indoor environment use case will develop such a HVAC system that will also 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. In that use case the goal is also to develop a healthy indoor environment in the buildings that can automatically adjust its conditions to better temperature, air quality, humidity, etc. using the results analyzed by the environmental data. The Rehabilitation decision support use case 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. The Indoor air quality improvement at school use case will combine the data from indoor air quality sensors and wearables with users' personal feedback and environmental sensing information to provide a healthier living environment for elderly people. The idea is to learn from sensor data by means of combining different data streams and applying environmental models, machine learning, data mining and supported by big data ICT.

The state-of -the-art description together with use case requirements will enable the implementing ESTABLISH system, and starting the project work. Next, the pilots will be planned more in detail and pilots will be implemented. This document will be updated during the project when necessary. Technical specifications will be described more in detail in the next deliverable of WP2.

References

3GPP website, www.3gpp.org (Accessed 29 May 2017)

Al Horr, Y., Arif, M., Kaushik, A., Mazroei, A., Katafygiotou, M. and Elsarrag, E. (2016). “Occupant Productivity and Office Indoor Environment Quality: A Review of the Literature.” Building and Environment 105: 369–89. doi:10.1016/j.buildenv.2016.06.001.

Alberdi, A., Aztiria, A. and Basarab, A. (2016). “Towards an Automatic Early Stress Recognition System for Office Environments Based on Multimodal Measurements: A Review.” Journal of Biomedical Informatics 59: 49–75. doi:10.1016/j.jbi.2015.11.007.

Anda, C., Fourie, P. and Erath, A. (2016). Transport Modelling in the Age of Big Data. Work Report. Future Cities Laboratory, Jun 2016.

Antoniou, C. (2004). “On-line calibration for Dynamic Traffic Assignment”. Ph Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 2004.

Antoniou, C., Ben-Akiva, M. and Koutsopoulos, H. N. (2005). “On-Line Calibration of Traffic Prediction Models”. Transportation Research Record 1934, 2005, pp. 235-245.

Antoniou, C., Ben-Akiva, M., Bierlaire, M., and Mishalani, R. (1997). “Demand simulation for dynamic traffic assignment”. In Proceedings of the 8th IFAC Symposium on Transportation Systems, Chania, Greece, 1997.

Ashok, K. and Ben-Akiva, M. (1993). “Dynamic O-D matrix estimation and prediction for real-time traffic management systems”. Transportation and Traffic Theory, pp. 465- 484. Elsevier Science Publishing 1993.

Ashok, K. and Ben-Akiva, M. (2000). “Alternative approaches for real-time estimation and prediction of timedependent origindestination flows”. Transportation Science, 2000, vol. 34, no.1, pp. 21-36.

Auradkar, A., Botev,C., Das, S., et al., “Data infrastructure at Linkedin,” in 28th International Conference on Data Engineering, Washington, USA, 2012, pp. 1370–1381. doi:

10.1109/ICDE.2012.147.

Avasthi, S. and Dwivedi, A. (2013). “Predeiction of Mobile User Behavior Using Clustering”. International Journal of Scientific and Research Publications, Volume 3, Issue 2, February 2013.

Axhausen, K. W. (2006). “Social factors in future travel: A qualitative assessment”. IEEE Proceedings Intelligent Transport Systems, 2006, 153 (2): 156–166.

Axhausen, K. W. and Frei, A. (2007). “Size and structure of social network geographies”, Arbeitsbericht Verkehrs und Raumplanung, 2007, No. 439, ETH Zurich.

Bañon Blazquez L. and Bevía García J. F. (2000). “Manual de Carreteras.Universidad de Alicante”. Departamento de Ingeniería de la Construcción, Obras Públicas e Infraestructura Urbana, 2000.

Ben-Akiva, M., Bierlaire, M., Koutsopoulos, H. N., and Mishalani, R. (2002). “Real-time simulation of traffic demand-supply interactions within DynaMIT”. In Gendreau, M. and Marcotte, P., editors, Transportation and network analysis: current trends, pp. 19–36. Kluwer Academic Publishers, 2002.

van den Berg, P. E. W., Arentze, T. A. and Timmermans, H. J. P. (2009). “Size and composition of ego-centered social networks and their effect on travel distance and contact frequency”. Transportation Research Record: Journal of the Transportation Research Board , 2009, 2135, 1-9.

van den Berg, P. and Timmermans, H. (2015). “A multi-level path analysis of social networks and social interaction in the neighbourhood”. Region, 2015, 2 (1), 55-66.

Bierlaire, M. and Crittin, F. (2004). “An efficient algorithm for real-time estimation and prediction of dynamic OD tables”. Operations Res, vol. 52, no.1, 2004.

Bluetooth. “A look at the basics of Bluetooth technology” n.d. http://www.bluetooth.com/Pages/Basics.aspx, (Accessed 23 May 2017)

Bodys-Cupak, I., Majda, A., Zalewska-Puchała, J., and Kamińska, A. (2016). “The Impact of a Sense of Self-Efficacy on the Level of Stress and the Ways of Coping with Difficult Situations in Polish Nursing Students.” Nurse Education Today 45: 102–7. doi:10.1016/j.nedt.2016.07.004.

Braekers, K., Ramaekers, K. and I. Van Nieuwenhuyse (2015). “The Vehicle Routing Problem: State of the Art Classification and Review”, 2015.

Brodesser, M. (2013). “Multi-Modal Route Planning”. Master Thesis, University of Freiburg Faculty of Engineering Department for Computer-Science, 2013.

Buchin, K., Cabello, S., Gudmundsson, J., Löffler, M., Luo, J., Rote, G., Silveira, R. I., Speckmann, B. and Wolle, T. (2010). “Finding the Most Relevant Fragments in Networks“. Journal of Graph Algorithms and Application, 2010, 14(2): 307–336.

Buchin, K., Buchin, M., Gudmundsson, J., Löffler, M. and Luo, J. (2011). “Detecting Commuting Patterns by Clustering Subtrajectories”, International Journal of Computational Geometry and Applications, special issue on 19th International Symposium on Algorithms and Computation (ISAAC), 2011, Vol. 21(3), pp. 253-282.

Buchin, M., Kruckenberg, H. and Kölzsch, A. (2012). “Segmenting Trajectories based on Movement States”, In Proc. 15th International Symposium on Spatial Data Handling (SDH). Advances in Spatial Data Handling: Geospatial Dynamics, Geosimulation and Exploratory Visualization, 2012, pp. 15-25.

Burghout, W., "Hybrid Microscopic-Mesoscopic Traffic Simulation Modelling". PhD thesis, 2004.

Carrasco, J.A., Hogan, B., Wellman, B., Miller, J. (2008). “Collecting social network data to study social activity-travel behavior: an egocentric approach”. Environment and Planning B: Planning and Design, 2008, 35 (6), 961-980.

Cheng, S.-W., Garlan, D. and Schmerl, B. (2012). “Stitch: A language for architecture-based self-adaptation,” Journal of Systems and Software, vol. 85, no. 12, pp. 1–38, 2012.

Cipriani, E. (2012). “Traffic state estimation based on data fusion techniques”. In IEEE Conference on Intelligent Transportation Systems, 15th IEEE ITSC, 2012, p. 147-148.

Csaji, B. C. and Monostori, L. (2008). "Adaptive Stochastic Resource Control: A Machine Learning Approach", Journal of Artificial Intelligence Research, 2008, Volume 32, pp. 453-486.

Damiani, M.L., Silvestri, C. and Bertino, E. (2011). “Fine-grained cloaking of sensitive positions in location sharing applications”. IEEE PERVASIVE COMPUTING, 2011, vol. 10 (4), pp. 64-72 (ISSN 1536-1268).

Dean, J., & Ghemawat, S. (2008). “MapReduce: simplified data processing on large clusters”. Communications of the ACM, 2008, 51(1), 107-113.

Delen, D. and Demirkan, H. (2013). “Data, information and analytics as services”, Decision Support Services, 2013, 55(1), pp. 359-363.

Demšar, U. and Virrantaus, K. (2010). “Space-time density of trajectories: exploring spatio-temporal patterns in movement data”. Intl. Journal of Geographical Information Science, 2010, 24(10): 1527–1542.

Distefano, S., Merlino, G. and Puliafito, A. (2012). "Sensing and actuation as a service: A new development for clouds." Network Computing and Applications (NCA), 2012 11th IEEE International Symposium on. IEEE.

Dobkin, D.M and Aboussouan, B. (2009). "Low power Wi-Fi (IEEE 802.11) for IP smart objects", GainSpan Corporation n.d. http://www.gainspan.com/docs2/Low_Power_Wi-Fi_for_Smart_IP_Objects_WP_cmp.pdf (Accessed 23 May 2017)

Duong, T. and Tran, D. (2015). “A Fusion Of Data Mining Techniques For Predicting Movement Of Mobile Users”. Journal of Communications and Networks, Volume: 17, Issue: 6, Dec. 2015, pp. 568-581.

European Commission (2007). The SCHER opinion on risk assessment on indoor air quality - European Commission, DG SANCO, Scientific Committee on Health and Environmental Risks, 2007.

European Environment Agency (EEA), http://www.eea.europa.eu/.

Feng, D., Jiang, C., Lim, G., Cimini, L. J., Feng, G., & Li, G. Y. (2013). "A survey of energy-efficient wireless communications". IEEE Communications Surveys & Tutorials, 15(1), 167-178.

Ferreira, P. Specification and Implementation of an Artificial Transport System. Engineering, 2008.

Firstbeat Technologies Ltd. 2014. “White Paper: Stress and Recovery Analysis Method Based on 24-Hour Heart Rate Variability.” https://assets.firstbeat.com/firstbeat/uploads/2015/10/Stress-and-recovery_white-paper_20145.pdf.

Francois, J. M. (2007). “Performing and Making Use of Mobility Prediction”. UNIVERSIT´E DE LI`EGE. Thesis, 2007.

Gambs, S. and Killijian, M. O. (2012). “Next Place Prediction using Mobility Markov Chains”. MPM’12, April 10, 2012, Bern, Switzerland.

Gavalas, D., Konstantopoulos, C., Mastakas, K. and Pantziou, G. (2014). ”A Survey on Algorithmic Approaches for Solving Tourist Trip Design Problems”. Journal of Heuristics, June 2014, Volume 20, Issue 3, pp. 291-328.

Geisberger, R. (2011). “Advanced Route Planning in Transportation Networks”. Ph. Dissertation. Karlsruher Instituts für Technologie, 2011.

Gerostathopoulos, I., Bures, T., Hnetynka, P., Keznikl, J., Kit, M., Plasil, F. and Plouzeau, N. (2015). “Self-Adaptation in Cyber-Physical Systems: from System Goals to Architecture Configurations,” Department of Distributed and Dependable Systems, D3S-TR-2015-02, D3S-TR-2015-02, Apr. 2015.

Gjoreski, M., Gjoreski, H., Luštrek, M. and Gams, M. (2016). “Continuous Stress Detection Using a Wrist Device – In Laboratory and Real Life.” In UBICOMP/ISWC ’16, 1185–93. doi:10.1145/2968219.2968306.

Gomez, C., & Paradells, J. (2010). "Wireless home automation networks: A survey of architectures and technologies". IEEE Communications Magazine, 48(6), 92-101.

Greenfeld J.S. (2002). “Matching GPS Observations to Locations on a Digital Map”, 2002, Transportation Research Board, Washington DC. http://www.docin.com/p-201213365.html [available Sep. 2014].

Gustafsson, F., Gunnarsson, F., Bergman, N., Forssell, U., Jansson, J., Karlsson, R., Nordlund, P. (2002). “Particle filters for positioning, navigation, and tracking”, IEEE Transactions on Signal Processing 50, 425–435, 2002.

Hoogendoorn, S.P., Bovy, P. H. L. State-of-the-art of vehicular traffic flow modelling. Proceedings of the I MECH E Part I Journal of Systems & Control in Engineer, 215(4):283–303, August 2001.

Huebscher, M. C. and McCann, J. A. (2008). “A Survey of Autonomic Computing - Degrees, Models, and Applications,” ACM Comput. Surv., vol. 40, no. 3, pp. 7:1–7:28, Aug. 2008.

Ibbotson, J. et al. (2010). "Sensors as a service oriented architecture: Middleware for sensor networks." Intelligent Environments (IE), 2010 Sixth International Conference on IEEE.

Iivari, A., Väisänen, T., Ben Alaya, M., Riipinen, T., & Monteil, T. (2014). "Harnessing XMPP for Machine-to-Machine Communications & Pervasive Applications". Journal of Communications Software & Systems, 10(3).

Jones, A.P. 1999. “Indoor Air Quality and Health.” Atmospheric Environment 33: 4535–64.

Järv, O., Ahas, R., Saluveer, E., Derudder, B. and Witlox, F. (2012). “Mobile Phones in a Traffic Flow: A Geographical Perspective to Evening Rush Hour Traffic”, PLoS One, 11/2012, Volume 7, Issue 11.

Kaur, N. and Monga, S. (2014). "Comparisons of wired and wireless networks: A review. International Journal of Advanced Engineering Technology", V(II).

Kokkinogenis, Z., Sanchez, L., Rossetti, R., Gabriel, J. "Towards the next-generation traffic simulation tools: a first evaluation Future Urban Transport., DSIE’11 - 6th Doctoral Symposium on Informatics Engineering, Porto, Portugal, January 2011

Koch, W. (2010). “On Bayesian Tracking and Data Fusion: A Tutorial Introduction with Examples”. IEEE A&E SYSTEMS MAGAZINE VOL. 25, NO. 7, JULY 2010.

Kraft, S., Casale,G., Jula, A., et al., “WIQ: work-intensive query scheduling for in-memory database systems,” in IEEE 5th International Conference on Cloud Computing, Honolulu, USA, 2012, pp. 33–40. doi: 10.1109/CLOUD.2012.120.

Lamsweerde, A.V. (2008). “Requirements Engineering: from Craft to Discipline,” in Proc. of FSE ’08, 2008, pp. 238–249.

Lan, L., Wargocki, P., Wyon, D. P. and Lian, Z. (2011). “Effects of Thermal Discomfort in an Office on Perceived Air Quality, SBS Symptoms, Physiological Responses, and Human Performance.” Indoor Air 21 (5). Blackwell Publishing Ltd: 376–90. doi:10.1111/j.1600-0668.2011.00714.x.

Latvakoski, J., Iivari, A., Vitic, P., Jubeh, B., Alaya, M. B., Monteil, T. & Kellil, M. (2014). "A Survey on M2M Service Networks". Computers, 3(4), 130-173.

Laube, P. and Purves, R. (2011). “How fast is a cow? Cross-Scale Analysis of Movement Data”. Transactions in GIS 15 (3), 2011, pp. 401–418.

Liu, W., Zhong, W. and Wargocki, P. (2017). “Performance, Acute Health Symptoms and Physiological Responses during Exposure to High Air Temperature and Carbon Dioxide Concentration.” Building and Environment 114: 96–105. doi:10.1016/j.buildenv.2016.12.020.

LoRa Alliance www-pages n.d. https://www.lora-alliance.org/what-is-lora/technology (Accessed 23 May 2017).

Ma, H. D. (2011). "Internet of things: Objectives and scientific challenges". Journal of Computer science and Technology, 26(6), 919-924.

Maddalena, R., Mendell, M. J., Eliseeva, K., Chan, W. R., Sullivan, D. P., Russell, M., Satish, U. and Fisk, W. J. (2015). “Effects of Ventilation Rate per Person and per Floor Area on Perceived Air Quality, Sick Building Syndrome Symptoms, and Decision-Making.” Indoor Air 25 (4): 362–70. doi:10.1111/ina.12149.

Marz, N. and Warren, J. (2013). “Big Data: Principles and best practices of scalable real-time data systems”. Manning Publications, 2013.

Mirkovic, M., Ćulibrk, D., Papadopoulos, S., Zigkolis, C., Kompatsiaris, Y., McArdle, G. and Crnojevic, V. (2011). “A Comparative Study of Spatial, Temporal and Content-based Patterns Emerging in YouTube and Flickr”. The International Conference on Computational Aspects of Social Networks, Salamanca, Spain, October 19 -21, 2011.

Mukhopadhyay, S. C. (2015). “Wearable Sensors for Human Activity Monitoring: A Review.” IEEE Sensors Journal 15 (3): 1321–30. doi:10.1109/JSEN.2014.2370945.

NearFieldCommunication.org n.d. http://nearfieldcommunication.org/ (Accessed 23 May 2017).

Neumeyer, L., Robbins, B., Nair, A., et al., “S4: distributed stream computing platform,” in IEEE International Conference on Data Mining Workshops, Sydney, Australia, 2010, pp. 170–177. doi: 10.1109/ICDMW.2010.172.

Ochieng W.Y., Quddus M. and Noland R.B. (2003). “Map-Matching in Complex Urban Road Networks”, Brazilian Journal of Cartography, 2003, 55 (2), pp. 1-14.

Ohnmacht, T., Schaffner, D., Weibel, C. and Schad, H. (2017). “Rethinking social psychology and intervention design: A model of energy savings and human behaviour”. Energy Research & Social Science, 2017, C(26), 40-56.

Olabarrieta, I., Torre, A. I. and Molinete, B. (2014). Autopistas Inteligentes.

Papageorgiou, M., Wang, Y. and Messmer A. (2007). “Real-Time Freeway Traffic State Estimation Based on Extended Kalman Filter: A Case Study”. Journal Transportation Science. Volume 41 Issue 2, May 2007, pp. 167-181.

Park, M., Hong, J. H. and Cho, S. B. (2006). Two-Stage User Mobility Modeling for Intention Prediction for Location-

Based Services. E. Corchado et al. (Eds.): IDEAL 2006, LNCS 4224, pp. 538 – 545, 2006. Springer-Verlag Berlin Heidelberg.

Pelekis, N. and Theodoridis, Y. (2014). “Mobility Data Management and Exploration”, Springer, 2014.

Piwek, L., Ellis, D. A., Andrews, S., Joinson, A., Maisel, WH and Ho, J. (2016). “The Rise of Consumer Health Wearables: Promises and Barriers.” PLOS Medicine 13 (2). MIT Press: e1001953. doi:10.1371/journal.pmed.1001953.

Persily, A. (2015). “Challenges in developing ventilation and indoor air quality standards: The story of ASHRAE Standard 62”, Building and Environment, Volume 91, September 2015, Pages 61-69, http://dx.doi.org/10.1016/j.buildenv.2015.02.026.

Postscapes (2014), “Internet of Things Technologies”, available at: http://postscapes.com/internet-of-things-technologies (accessed May 24, 2017)

Prabakar, J., Mohan, V. and Ravisankar, K. (2015). “Evaluation of Low Cost Particulate Matter Sensor for Indoor Air Quality Measurement”, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 4, Issue 2, February 2015, DOI: 10.15680/IJIRSET.2015.0402076 366.

Prasad, P.S. and Agrawal (2010). “Movement Prediction in Wireless Networks Using Mobility Traces”. IEEE Communications Society subject matter experts for publication in the IEEE CCNC 2010 proceedings.

Pyo J., Shin D. and Sung T. (2001). “Development of a map-matching method using the multiple hypothesis technique”, IEEE Proceedings on Intelligent Transportation Systems, 23–27, 2001.

Quddus, M.A., Ochieng, W.Y. and Noland, R.B. (2007). “Current map-matching Algorithms for Transport Applications: State-of-the art and future Research Directions”, Transportation Research Part C 15, 2007, pp. 312-328.

Renso, C., Spaccapietra, S. and Zimányi, E. (2013). “Mobility Data: Modeling, Management, and Understanding”. Cambridge University Press, 2013, 368 p.

Rodgers, M. M., Pai, W. M. and Conroy, R. S. (2015). “Recent Advances in Wearable Sensors for Health Monitoring.” IEEE Sensors Journal 15 (6): 3119–26. doi:10.1109/JSEN.2014.2357257.

Sandulescu, V., Andrews, S., Ellis, D., Bellotto, N. and Martínez Mozos, O. (2015). “Stress Detection Using Wearable Physiological Sensors.” In , 526–32. Springer, Cham. doi:10.1007/978-3-319-18914-7_55.

Schultes, D. (2018). “Route Planning in Road Networks”. Ph. Dissertation. Karlsruher Instituts für Technologie, 2008.

Shany, T., Redmond, S. J., Narayanan, M. R. and Lovell, N. H. (2012). “Sensors-Based Wearable Systems for Monitoring of Human Movement and Falls.” IEEE Sensors Journal 12 (3). IEEE: 658–70. doi:10.1109/JSEN.2011.2146246.

Shelby, Z., & Bormann, C. (2011). 6LoWPAN: The wireless embedded Internet (Vol. 43). John Wiley & Sons.

SigFox, Sigfox technology overview: https://www.sigfox.com/en (Accessed 24 May 2017)

Sinha, R.S., Wei, Y., & Hwang, S-H, (2017). "A survey on LPWA technology: LoRa and NB-IoT, ICT Express", Volume 3, Issue 1, March 2017, Pages 14-21, ISSN 2405-9595, https://doi.org/10.1016/j.icte.2017.03.004.

Sisäilmayhdistys, http://www.sisailmayhdistys.fi/ (in Finnish).

Sommer, C., Dietrich, I., Dressler, F.. Realistic Simulation of Network Protocols in VANET Scenarios. 2007 Mobile Networking for Vehicular Environments, pages 139–143, 2007

Soriguera, F. (2011). “Highway travel time estimation with data fusion”. Ph. Thesis. Cátedra Abertis de Gestión de Infraestructuras del Transporte. Universitat Politècnica de Catalunya, 2011.

Stein, P.K., and Kleiger, R.E. (1999). “Insights from the Study of Heart Rate Variability.” Annual Review of Medicine 50 (1): 249–61.

Stenudd, S. (2010), "Using machine learning in the adaptive control of a smart environment”, VTT Publications 751, 2010.

Su, C., Wu, D., Xie, X.Z. and Wu, Q. (2016). “Temporal Markov Chain Location Prediction”. 2016 International Conference on Applied Mechanics, Mechanical and Materials Engineering (AMMME 2016).

Sundell, J., Levin, H., Nazaroff, W. W., Cain, W. S., Fisk, W. J., Grimsrud, D. T., Gyntelberg, F. et al. (2011). “Ventilation Rates and Health: Multidisciplinary Review of the Scientific Literature.” Indoor Air 21 (3). Blackwell Publishing Ltd: 191–204. doi:10.1111/j.1600-0668.2010.00703.x.

Tao, G. (2014), “Multivariable adaptive control: A survey”, Automatica, Volume 50, Issue 11, 2014, Pages 2737-2764, ISSN 0005-1098, http://dx.doi.org/10.1016/j.automatica.2014.10.015.

Timoteo, I. J. P. M, Araujo, M.R., Rossetti, R.J.F, Oliveira, E.C. "Trasmapi: An api oriented towards multi-agent systems real-time interaction with multiple traffic simulators". In Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on, pages

1183 –1188, sept. 2010.

Tham, K. W. (2016). “Indoor Air Quality and Its Effects on humans—A Review of Challenges and Developments in the Last 30 Years.” Energy and Buildings 130: 637–50. doi:10.1016/j.enbuild.2016.08.071.

TSI (2013), “Indoor air quality handbook- a practical guide to indoor air quality investigations”, TSI, 2013.

Wang, R., Blackburn, G., Desai, M., Phelan, D., Gillinov, L., Houghtaling, P. and Gillinov, M. (2017). “Accuracy of Wrist-Worn Heart Rate Monitors.” JAMA Cardiology 2 (1). American Medical Association: 104. doi:10.1001/jamacardio.2016.3340.

Wang, Y. M. and Papageorgiou (2005). “Real-time freeway traffic state estimation based on extended Kalman filter: a general approach”. Transportation Research Part B: Methodological, 2005, 39(2): 141-167.

Want, R. (2006). "An introduction to RFID technology". IEEE Pervasive Computing, 5(1), 25-33.

Wargocki, P. and Wyon, D. P. (2017). “Ten Questions Concerning Thermal and Indoor Air Quality Effects on the Performance of Office Work and Schoolwork.” Building and Environment 112: 359–66. doi:10.1016/j.buildenv.2016.11.020.

Wargocki, P., Wyon, D. P., Baik, Y. K., Clausen, G. and Fanger, P. O. (1999). “Perceived Air Quality, Sick Building Syndrome (SBS) Symptoms and Productivity in an Office with Two Different Pollution Loads.” Indoor Air 9 (3). Munksgaard International Publishers: 165–79. doi:10.1111/j.1600-0668.1999.t01-1-00003.x.

Weightless, http://www.weightless.org/about/what-is-weightless (Accessed 24 May 2017)

Wi-Fi Alliance n.d. http://www.wi-fi.org/ (Accessed 23 May 2017).

Wikipedia. 2017. “Sick Building Syndrome.” https://en.wikipedia.org/wiki/Sick_building_syndrome.

WiMax.com, What is WiMAX? n.d. http://www.wimax.com/general/what-is-wimax (Accessed 23 May 2017).

WHO (2010). WHO guidelines for indoor air quality: selected pollutants, 2010.

Yen, S. (2009). “NoSQL is a horseless carriage”. November 2009. http://dl.getdropbox.com/u/2075876/nosql-steve-yen.pdf

van der Zijpp, N. J. (1997). “Dynamic OD-Matrix Estimation from Traffic Counts and Automated Vehicle Identification Data”. Transportation Research Record, 1997, no. 607, pp. 87–94.

Zhou, X. and Mahmassani, H. S. (2006). “Dynamic origin destination demand estimation using automatic vehicle identification data”. IEEE Transactions on Intelligent Transportation Systems, 2006, vol. 7, no. 1, pp. 105-114.

ZigBee website n.d. www.zigbee.org/ (Accessed 23 May 2017).

Züger, M. and Fritz, T. (2015). “Interruptibility of Software Developers and Its Prediction Using Psycho-Physiological Sensors.” In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI ’15, 2981–90. doi:10.1145/2702123.2702593.

Appendixes

Appendix 1A. Optimized City and Mobility Planning, technical requirements

	Discription
Devices, sensor management (WP3 System architecture)	
Other data sources	Open data from Valencia City Council - Valencia Smart City Platform (http://gobiernoabierto.valencia.es/en)
Data (WP4 Connecting and managing sensors)	
Data sources	
Storage	Cloud platform
Recording of data	Air pollution from weather stations Traffic data in real time Static data of public transport (buses) Bike lanes Public and private parkings Information about regulated parking service
Monitoring	Air pollution from weather stations Traffic data in real time Static data of public transport (buses) Bike lanes Public and private parkings Information about regulated parking service
Gateways	FI-WARE
Integration	Open Data from Valencia Smart City Platform analyzed and integrated in the ESTABLISH platform
Analytics (WP5 Data analytics and adaptive control)	
	Data fusion based on big data and deep learning techniques in order to provide recommendations for mobility and alerts about pollution in real time
	Data analysis with algorithms using geographical aware complex event processing (geoCEP) in realtime situations.
Privacy and Security	
Application	
User needs	Applications providing planning services and mobility information considering all relevant information about traffic conditions and pollution
Front end	Mobile application with multimodal route planning and alerts/recommendations about pollution. Web application with dashboard and traffic simulation platform
Back end	Algorithms for ecological route optimization Dashboard Traffic Simulation Platform
Pilot Implementation	
Implementation	In progress
Integration	To be done
Testing	To be done

Appendix 1B. Optimized City and Mobility Planning, use case requirements

PILOT	Optimized City and Mobility Planning
COUNTRY	Spain
DESCRIPTION	The pilot 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..
OWNER, CONTACT PERSON (in the project)	Ismael Torres (Prodevelop)

	Description
Use case preparation	
Brainstorming	Kick-off of spanish partership (20/02/2017).
User scenario development	Plan for scenario writing D2.1, preliminary of work structure and invovement of partners, candidate technologies, etc.
User experience planning	User experience viewpoints in the pilot / how / how often... Aproach based on an iterative process with mockups, to be refined once the use case definition is concluded
Contacting the users, recruitment	User community building. Check use case with final users (City interface and responsables (mobility, infrastrcture, smart city and innovation)). Revent communities, associations and relventa users (citizen view)
Informing about the pilot	Contrast meetings with Valencia City responsables
Business modeling	At this stage identification of relevant stakeholders and partners. Business Model Canvas to be developed duirng the project lifecycle
Evaluation criterias	When we have succeeded. Brainstorming around the initial proposal (KPIs and relevant data acquisition processes)
Use case specification	
Use case definition	Definition of the spanish use case according the common template defined for the project.
Business modeling	At this stage identification of relevant stakeholders and partners. Business Model Canvas to be developed duirng the project lifecycle
Feedback from stakeholders	Meetings with Las Naves-InnDEA, entity of the Valencia City for innovation issues (representants for Smart City and mobility innitatives). To prepare questionnaires and groups dynamics to refine and evaluate in an incemrental way the different developments.
Feedback from advisory board	To define and involve reference compies and organizations to complete the spanish chapter. To prepare: formal invitation, questionnaires, interviews, group discussions, co-creation tools...
Technical specifications	See Technical specifications sheet

Use case design	
Concept visualization	Adopting a lean or incremental development lifecycle, different tools and methods will be used to align work among partners under a common picture (paper prototyping, mock-ups…)....
Feedback from stakeholders	Concept evaluation, iterative meetings with relvant staksholders (city and final users associations, if applies)
Feedback from advisory board	Concept evaluation
Implementation	
Contacting the users	A common plan to contact and involve users (How, how often, what to inform…). With the City, a communication and coordination scheme have been established during the first weeks of the project (see KoM minutes); with other users, this plan will be defined once user engagement is running and a clear concept of the pilot available
Review meetings	Meetings (F2F or teleconference) involving all the pilot partners are expected each two weeks or monthly at latest. Minutes will be prepared and agred among partners.
Detailed architectural specifications	Detailed architectural specifications
Detailed architectural design	Detailed architectural design
Integration plan	Integration plan
validation plan	validation plan
During the pilot	
Contacting the users	According the common plan to contact and involve users (how, how often, what to inform…), involve the city on the pilot expoitation, KPI monitoring and evoluation, risks and integration on the public strategy. Contact channels with users will keep alive during the piloting stage, by means of questionaires and working gruops dynamics if possible.
User experience	Selection of tools for Questionnaires, interviews, group discussions, co-creation). Involvement of users.
After the pilot	
Evaluation of the pilot	Evalaution with final users: KPIs final evaluation, sucess of project according just defined criteria
Contacting the users	By using previous tools, acquire feedback of users
Conclusions and recommendations	Extract conclusions, best practices, guidelines from user, stakeholders and advisory group feedback for next steps of the consortium and European community.
Feedback from stakeholders	Specific working meetings to analyze business and co-deisng nof business oportunities, technology evalaution and new lines of work opening
Feedback from advisory board	Evalaution with final users: KPIs final evaluation, sucess of project according just defined criteria and bsuienss opportunities.
Feasibility check	Assessing the feasibility of the concept to go to market and defining the necessary steps
Business model update	Finalisation of the business model canvas

Appendix 2A. Smart HVAC systems that ensure a healthy indoor environment, technical requirements

	Discription
	Devices, sensor management (WP3 System architecture)
Sensor 1	Temperature sensor(s)
Sensor 2	Humidity sensor(s)
Sensor 3	CO2 sensor(s)
Sensor 5	PM sensor(s) (optional)
Sensor 6	Radon gas sensor(s) (optional)
Sensor 7	Iluminance sensor (s)
Sensor 8	Noise sensor(s)
Acuator 1	Window opener
Acuator 2	Local AHU with heat recovery
Acuator 3	Thermostat / Heater valves
Data source	Online weather data (e.g. CHMI)
	Data (WP4 Connecting and managing sensors)
Storage	MS Azure cloud platform
Recording of data	Location-fixed indoor conditions data (temperature, humidity, CO2, pressure difference wrt outdoors) every 5 minutes.
Recording of data	Outdoor conditions data; every 5 minutes
Recording of data	Indoor air quality data from fixed sensors every 5 minutes or according to sensor specification (to be defined)
Recording of data	Physiological sensors; maximum rate available by sensors for one day operation between charging.
Recording of data	User-defined labels; user initiated (any time) and scheduled (three times during working day; morning, midday, when leaving work). Option for backend-originated querying of user label based on observed physiological data.
Recording of data	Air purifier; filtering efficiency (e.g., power vs airflow and/or purification efficiency). Optional: Incoming air quality (if not available via the external indoor air quality sensors)
Monitoring	Visualizations updated at least once a day for participants.
Monitoring	Air purifier filter status; daily

Gateways	Location-fixed indoor sensors connect via IMA gateways and/or vendor specific selected cloud IoT Hub. Mobile phone may be used as gateway for pilot participants' physiological sensors.
Connectivity	Sensor <=> gateway via LoRa and/or BTLE Gateway <=> cloud via 3G/4G cellular and/or WiFi
Integration	IMA indoor conditions conditions data are made available to project consortium.
	Analytics (WP5 Data analytics and adaptive control)
Mathematical model	A mathematical representation of the building will be used to control the HVAC system in the pilot. Model wiil be able to use measered data to predict optimal parameters of the HVAC system in the future.
	Privacy and Security
	Privacy and security provisions will be defined.
Application	
User needs	Interviews / questionnaires / focus group discussions to understand users needs.
Front end	Sensor data collection; visualizations of collected and predicted data. Visualisation of acuators settings.
Back end	Data aggregation for analytics, visualization and prediciton. Connecting the measured date with mathematical model.
Front end / Back end (division TBD)	Adjustment of acuators based on observed environmental conditions and predicted conditions.
	Pilot Implementation
Implementation	what, when, who to cantact, how, who…
Testing	Phase 1: At IMA premises. Phase 2: TBD when stakeholders interviewed

Appendix 2B. Smart HVAC systems that ensure a healthy indoor environment, use case requirements

Use case requirements and technical specifications	
PILOT	Smart HVAC systems that ensure a healthy indoor environment
COUNTRY	Czech republic, Korea
DESCRIPTION	The pilot will show up defined technical as well as formal aspects of Czech and Korean use cases. Partners will develop HW and SW components of the use cases - sensors, gateways, coordination platfor, SW services, applications including data interpretation interfaces. Determined stakeholders will be familiar with functionalities and expected results. Use cases will be deployed at selected test sites in Czech and Korea. The focus of the pilot is to assess the impact of smart HVAC technologies on health status improvement.
OWNER, CONTACT PERSON (in the project)	Jiri Havlik (IMA)
	Description
	Use case preparation
Brainstorming	Discussing with partners involved.
User scenario development	Scenario writing for D2.1.
User experience planning	User experience viewpoints in the pilot / how / how often etc. will be planned together with use case developers - IMA, CU, DEKPROJEKT, ETRI, Coway.
Contacting the users, recruitment	The contact person will be contacted and a planning meeting will be arranged.
Informing about the pilot	Informative meetings will be arranged before, during and after the test period.
Business modeling	Business model canvas will be created when the pilot will be defined. The canvas will be discussed in the steering group meeting also.
Evaluation criterias	TBD
	Use case specification

Use case definition	The Czech pilot will develop such a HVAC system that will also 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. The Korean pilot will contribute to developing a healthy indoor environment in the buildings that can automatically adjust its conditions to better temperature, air quality, humidity, etc. using the results analyzed by the environmental data. It will achieve the solution to generate the well-being, safe, comfortable indoor spaces for the people, especially the senior and the children.
Business modeling	Business model canvas will be developed during the pilot.
Feedback from stakeholders	Questionnaires, interviews, group discussions, co-creation tools will be utilized to collect feedback. It will be determined together with testee how the feedback will be collected during the pilot.
Technical specifications	See the other sheet: Technical specifications
Use case design	
Concept visualization	Paper prototyping, mock-ups will be utilized. The real application will be developed iteratively based on the feedback got from the users.
Feedback from stakeholders	UC concept and design will be validated using stakeholders' feedback and with DEKPROJEKT support. The feedback will be assessed by US developers.
Implementation	
Contacting the users	The contact person will be contacted and a planning meeting will be arranged. All testees will be informed about pilot.
Review meetings	The stakeholders will be informed at the meeting before the pilot. At the end of the pilot, a feedback event / group discussion meetings will be arranged.
During the pilot	
Contacting the users	TBD
User experience	Questionnaires, interviews, group discussions, co-creation tools will be utilized to collect feedback. It will be determined together with testee how the feedback will be collected during the pilot.
Analytics	Differences between predicted and measured indoor air quality will be evaluated during the pilot. Measures to maximally closing the gap will be developed.
After the pilot	
Evaluation of the pilot	The pilot will be evaluated with the stakeholders by the criterias defined before the pilot.

Contacting the users	The results of the pilot will be shared with the users in a meeting / by email.
Conclusions and recommendations	The results of the pilot will be analysed and shared with the stakeholders in a meeting.
Feedback from stakeholders	Feedback from the stakeholders will be collected in a meeting. They are also able to comment and update the report of the results.
Feasibility check	Assessing the feasibility of the concept to go to market and defining the necessary steps.
Business model update	Finalisation of the business model canvas with the stakeholders.

Appendix 3A. Rehabilitation decision support, technical requirements

	Description
Devices, sensor management (WP3 System architecture)	
Sensor 1	Air Quality Sensors (posibil Aeroqual series 500): sulfur dioxide (SO2), nitrogen oxides (NOx), carbon monoxide (CO), ozone (O3), particulate matter (PM10 and PM2)
Sensor 2	Weather sensors (pressure, temperature, humidity, CO2)
Sensor 3	FitBit Bracelet (heart rate, pedometer, calories burned)
Data (WP4 Connecting and managing sensors)	
Data sources	
Storage	Cloud platform
Recording of data	sulfur dioxide (SO2), nitrogen oxides (NOx), carbon monoxide (CO), ozone (O3), particulate matter (PM10 and PM2)/ once at 3 minutes
Recording of data	pressure, temperature, humidity, CO2/once at 15 minutes
Recording of data	heart rate, pedometer, calories burned/once in a minute
Monitoring	sulfur dioxide (SO2), nitrogen oxides (NOx), carbon monoxide (CO), ozone (O3), particulate matter (PM10 and PM2)/ once at 3 minutes
Monitoring	pressure, temperature, humidity, CO2/once at 15 minutes
Monitoring	heart rate, pedometer, calories burned/once in a minute
Gateways	FIWARE (?) to be done after sensors purchasing
Connectivity	Sensors -M2M Platform -Decision Expert System - Mobile Apps (for notifications)
Integration	Weather and air quality data will be available for al the users of the integrating platform
Analytics (WP5 Data analytics and adaptive control)	
	to be done after sensors purchasing
Privacy and Security	
	Users personal data will be seen only for doctors and rehabilitation stuff
Application	
User needs	Maintaining their physical condition and performing exercises for physical activity in order to recovery/restoring of motor functionalities, both outdoors and indoors.
Front end	Personal data, medical data, sensors data entries
Back end	Air quality, weather, behavioral data and notifications from the expert system.
Pilot Implementation	
Implementation	Work in progress
Integration	to be done
Testing	to be done

Appendix 3B. Rehabilitation decision support, use case requirements

PILOT	Promoting independence of specific vulnerable groups
COUNTRY	Romania
DESCRIPTION	The pilot will study the use of variety of indoor sensors combined with inhabitant's personal feedback and environmental sensing information to provide a healthier living environment for elderly people in their homes or in a care facility.
OWNER, CONTACT PERSON (in the project)	Ana-Maria Baldea (Siveco)

	Description
Use case preparation	
Brainstorming	Done with **BEIA, Premium Wellness Institute** and **National Institute of Rehabilitation, Physical Medicine and Balneoclimatology Bucharest**
User scenario development	Done
User experience planning	User experience viewpoints in the pilot / how / how often…
Contacting the users, recruitment	The first pilot activities will take place with two volunteers groups: students from a high school from Bucharest and corporates from SIVECO. They intend to maintain their fitness and prevent obesity. After this we will choose people who have prescriptions for physical activity for the recovery/restoring of motor functionalities, both outdoors and indoors.
Business modeling	Business model canvas
Evaluation criterias	When we have succeeded

Use case specification	
Use case definition	The Romanian pilot will combine environmental sensor data with physiological and behavioural sensor data to empower patients in a rehabilitation clinic with decision support tools for behavioural choices and treatment options. The Romanian ESTABLISH project's goals are: - to monitor health parameters to constantly improve the health of the population through rehabilitation and spa care, specifically targeting the patient's functional aspect of integration in everyday life, environment and work. - to develop a decision support system and services based on the outdoors environment parameters and indoor location. - to reduce operations costs and improve quality of the services provided. We will use a first group of students from highschool and a second pilot group of adults. We will register their physical activities, the cardiac rithm, the burned calories, in order to find links between them and the air conditions.
Business modeling	Business model canvas
Feedback from stakeholders	Questionnaires, interviews, group discussions, co-creation tools…
Feedback from advisory board	Questionnaires, interviews, group discussions, co-creation tools…
Technical specifications	See the other sheet: Technical specifications
Use case design	
Concept visualization	Paper prototyping, mock-ups… (WYP)
Feedback from stakeholders	Concept evaluation
Feedback from advisory board	Concept evaluation
Implementation	
Contacting the users	How, how often, what to inform…
Review meetings	Who to involve, when, how often…
During the pilot	
Contacting the users	Users will receive notifications everytime is necessarily: overcoming their concentration level of sulfur dioxide (SO2), nitrogen oxides (NOx), carbon monoxide (CO), ozone (O3), particulate matter (PM10 and PM2)
User experience	Questionnaires, interviews, group discussions, co-creation tools…
After the pilot	
Evaluation of the pilot	to be done
Contacting the users	How, how often, what to inform…
Feasibility check	Assessing the feasibility of the concept to go to market and defining the necessary steps
Business model update	Finalisation of the business model canvas

Appendix 4A. Indoor air quality improvement at school, technical requirements

	Discription
Devices, sensor management (WP3 System architecture)	
Sensor 1	Indoor conditions monitoring; temperature, humidity, CO2, pressure difference. E.g., VTT TinyNode and/or Ouman WL-TEMP-RH
Sensor 2	Outdoor weather / air quality sensors at VTT; temperature, air pressure, humidity, CO2, precipitation, wind speed and direction, particulates, NO2, SO2, CO, O3
Sensor 3	Indoor VOC sensor(s) (portable)
Sensor 4	Indoor particulate sensor
Sensor 5	Activity sensors (e.g. FitBit)
Sensor 6	Bioharness (???) with heart rate, breathing rate
Sensor 7	Oxygen saturation and skin conductivity sensors
Sensor 8	(within air purifier) - optional: incoming air quality, purification efficiency
Sensor purchasing	Sensors 1, 2, 3, 4 are already available for VTT pre-pilot. Personal sensors and indoor conditions monitoring sensors need to be purchased.
Implementation	Implementations of the **TBD**
Laptops	Do we need laptops for the pilot, TBD.
Mobile phones	Phones are purchased for development purposes; no specific requirements (stock Android smartphone). In case separate phones are decided to be used by pilot participants, similar phones may be purchased. Otherwise, pilot participants may use their own (Android) phones.
Mobile application	Design and coding of the application. See Application part from this document.
Data (WP4 Connecting and managing sensors)	
Storage	MS Azure cloud platform
Recording of data	Location-fixed indoor conditions data (temperature, humidity, CO2, pressure difference wrt outdoors) every 5 minutes.
Recording of data	Outdoor conditions data; every 5 minutes
Recording of data	Indoor air quality data from movable sensors (VOC, particulates) every 5 minutes or according to sensor specification (to be defined)
Recording of data	Physiological sensors; maximum rate available by sensors for one day operation between charging.
Recording of data	User-defined labels; user initiated (any time) and scheduled (three times during working day; morning, midday, when leaving work). Option for backend-originated querying of user label based on observed physiological data.
Recording of data	Air purifier; filtering efficiency (e.g., power vs airflow and/or purification efficiency). Optional: Incoming air quality (if not available via the external indoor air quality sensors)
Monitoring	Visualizations updated at least once a day for participants.
Monitoring	Air purifier filter status; daily
Gateways	Location-fixed indoor sensors connect via VTT gateways and/or vendor specific (e.g., Ouman WL-Base) to MS Azure IoT Hub. Mobile phone may be used as gateway for pilot participants' physiological sensors.
Connectivity	Sensor <=> gateway via BTLE, ZigBee. Gateway <=> cloud via 3G/4G cellular and/or WiFi
Integration	VTT indoor conditions and outdoor conditions data are made available to project consortium.

Analytics (WP5 Data analytics and adaptive control)	
	Started during VTT pre-pilot; seek to find correlation between user-defined labels (feelings, symptoms, activity) and physiological measurements (e.g., heart rate, breathing rate). Find correlation between symptoms and spaces the participant visits. Seek to classify sources of indoor air quality problems by correlating symptoms (user-specified and physiological), indoor and outdoor air quality measurements, questionnaire
Privacy and Security	
	Selection of pilot (and VTT pre-pilot) participants according to ISEC questionnaire & policy; selection criteria or personal data won't be available to project members (even at ISEC). Collected physiological data will be anonymized by hashing the person identifier, and not storing the unhashed identifier in the data collection and processing system.
Application	
User needs	Interviews / questionnaires / focus group discussions to understand users needs. User needs the application for labelling data (feelings, symptoms, etc.) and in order to see their personal collected data (visualizations). If location cannot be determined otherwise, ask the user.
Front end	User data collection (labels); visualizations of collected physiological data. Positioning / location tracking.
Back end	Data aggregation for analytics and visualization. Option for querying users for data labels based on observed physiological reactions. Location / positioning (e.g., combine BT beacon data, WiFi identifiers, etc. to determine location)
Front end / Back end (division TBD)	Adjustment of air purifier operation based on observed environmental conditions, and determination + indication of need for filter cleaning & replacement.
Pilot Implementation	
Implementation	what, when, who to cantact, how, who…
Testing	At VTT pre-pilot

Appendix 4B. Indoor air quality improvement at school, use case requirements

PILOT	Indoor air quality improvement at school
COUNTRY	Finland
DESCRIPTION	The Indoor air quality improvement at school pilot will study the use of a variety of indoor sensors and wearables combined with users personal feedback and environmental sensing information to provide a healthier living environment for pupils, teachers, and other staff members.
OWNER, CONTACT PERSON (in the project)	Heidi Similä (VTT)

	Description
Use case preparation	
Brainstorming	Branstorming with researchers, discussing with companies involved.
User scenario development	Scenario writing for D2.1.
User experience planning	User experience viewpoints in the pilot / how / how often etc. will be planned together with researchers and pre-testee at VTT.
Contacting the users, recruitment	Contacting potential users from VTT by email. In the case of school, the contact person will be contacted and a planning meeting will be arranged.
Informing about the pilot	More information about time pilot will be shared to the users by email including all practical information, contact person of the project etc. Also, an information meetings will be arranged before the test period.
Business modeling	Business model canvas will be filled in by researchers when we are more familiar with the pilot. The canvas will be discussed in the steering group meeting also.
Evaluation criterias	TBD
Use case specification	
Use case definition	The Finnish pilot targets users of buildings with confirmed or suspected indoor air quality problems. The aim is to improve and simplify the detection of the type and severity of indoor air quality problems, location affected by them (e.g., classrooms or offices) by correlating questionnaire data, environmental data, user-provided data of their symptoms and mood, and physiological data collected of the pilot participants. Possibility to utilize the air purifiers in the pilot will be evaluated. Suitable participants are screened by a structured questionnaire (ISEC) and/or volunteers. The participants wear physiological sensors (bioharness, wrist-held device(s)) and report their feelings (symptoms, mood, tiredness etc.) on a mobile application. Environmental data is simultaneously collected of the locations and outside conditions. Location data is either user provided, derived on a coarse level by, e.g., GPS signal, and/or using indoor locationing/positioning.
Business modeling	Business model canvas will be filled in by researchers when we are more familiar with the pilot. The canvas will be discussed in the steering group meeting also.
Feedback from stakeholders	Questionnaires, interviews, group discussions, co-creation tools will be utilized to collect feedback. It will be determined together with testee how the feedback will be collected during the pilot.
Technical specifications	See the other sheet: Technical specifications

Use case design	
Concept visualization	Paper prototyping, mock-ups will be utilized e.g. to visualize the mobile application for VTT testee. The real application will be developed iteratively based on the feedback got from the users.
Feedback from stakeholders	Concept evaluation with companies involved and project partners in the meetings. Co-operation possibilities with other partners will be discussed during the project.
Implementation	
Contacting the users	The testee at VTT and at school will be informed by email about e.g. the time tables, their role in the pilot and sensor implementations beforehand. Also, ma f2f meeting will be arranged both at VTT and at school.
Review meetings	The testee at VTT and at school will be informed in the meeting before the pilot. At the end of the pilot, a feedback event / group discussion meetings will be arranged.
During the pilot	
Contacting the users	The contact person from the school will be contacted at least weekly to understand how everything is going in the pilot.
User experience	Questionnaires, interviews, group discussions, co-creation tools will be utilized to collect feedback. It will be determined together with testee how the feedback will be collected during the pilot. It will be clarified if it could be possible to get some feedback also from the pupils.
Analytics	Pursue to find correlation between user input, air quality monitoring and physiological sensors; determine need for additional measurements and/or adjust user app (add/change/remove questions, change questioning interval).
After the pilot	
Evaluation of the pilot	The pilot will be evaluated with the stakeholders by the criterias defined before the pilot.
Contacting the users	The results of the pilot will be shared with the users in a meeting.
Conclusions and recommendations	The results of the pilot will be analysed and shared with the stakeholders in a meeting.
Feedback from stakeholders	Feedback from the stakeholders will be collected in a meeting. They are also able to comment and update the report of the results.
Feasibility check	Assessing the feasibility of the concept to go to market and defining the necessary steps.
Business model update	Finalisation of the business model canvas with the stakeholders and researchers.