HAAGA-HELIA
University of Applied Sciences

Weather and Air Quality Data of Helsinki

Fairuz Ishrat Bhuiyan

Bachelor's Thesis
Degree Programme in
Business Information
Technology
2016

HAAGA-HELIA
University of Applied Sciences

Abstract

19.6.2016

Author(s)
Fairuz Ishrat Bhuiyan

Degree programme
Business Information Technology

Report/thesis title
Weather and Air Quality Data of Helsinki

Number of pages and appendix pages
47 + 1

The objective of this thesis was researching, analyzing and classifying the contents and quality of the weather and air quality data for the Cityzer project. The aim was to work with the partners of Cityzer: The Finnish Meteorological Institute, Helsinki Region Environmental Services Authority, Vaisala, Pegasor and Tampere University of Technology to understand how they collect data and where they are stored.

The thesis gives an insight on the weather and air quality situation in the Finnish capital, Helsinki, and studies the factors that affect weather and air quality, how weather and air quality can be measured and the current situation and problems of the air quality in Helsinki. It also gives a brief insight on Big Data and its architecture and also the Internet of Things.

The main research question is What kind of data is available for analysing weather and air quality in the Helsinki region?" and sub questions backing up the main question are the data provided by the Cityzer project partners, the weather and air quality data: the origin of the data, their contents and format, their storage place and how they were classified. The partners were interviewed to understand their role in the project to collect answers to the question mentioned above.

The conclusion is that most of the partners provided the data to other partners and third parties, such as FMI, HSY, TUT and Pegaosr. Vaisala manufactured devices and sold them to the other organizations, so that they could collect the data via the devices. FMI is also the founder and system orchestrator of the Cityzer project. Most of the air quality data were collected by monitoring stations and sensors, and the biggest concern was pollutants. The data contains detailed information on air quality emission sources and particle emissions and also metadata. The data are stored in hard drives, cloud services or in servers. In HSY's case, it sends the data to FMI through an air quality portal and FMI, in turn, publishes them online. The data are classified by different models, nowcasting, Finnish air quality index or weather parameters.

The mission of the Cityzer project is to develop new digital services and products to sup-port decision making processes related to weather and air quality in cities. This thesis will provide the partners with useful information on their data, which they can demonstrate to outside parties and it will help them when they launch the CityzerDEMO in Helsinki.

Keywords
weather, air quality, data, Cityzer, particulate matter, Helsinki

Terms and abbreviations

Cityzer	The national project, which this thesis is based on.
FMI	The Finnish Meteorological Institute
HSY	Helsinki Region Environmental Services Authority
TUT	Tampere University of Technology
Big Data	A term for large volume of structured and unstructured data that is so large, that it is difficult to process using traditional database and software techniques.
IoT	The Internet of Things - the network of physical objects embedded with electronics, software, sensors, and network connectivity that enables these objects to collect and exchange data.
Metadata	Data that describes other data.
PM2.5	Fine particulate matter
PM10	Thoracic particulate matter

Table of contents

1 Introduction

The topic of this thesis is “Weather and air quality data of Helsinki” and the main objective is researching, analyzing and classifying the contents of the weather and air quality data for the Cityzer project. The aim is to work with the weather companies and partners, such as Vaisala, Pegasor, and The Finnish Meteorological Institute and the client HSY, and collect data from their database and analyse it. This research is an analysis of the data form and its origin. The theory part of this thesis is divided into five parts: “Air Quality and Weather”, “Helsinki’s situation”, “Big Data”, “The Internet of Things” and “Data, Information, Knowledge and Value”. It will look deeper into the weather and air quality in cities, specifically in the Finnish capital, Helsinki, and study the factors that affect weather and air quality, how they can be measured, the currently situation and problems or the air quality in Helsinki and how they can be measured. It will also look into Big Data and its architecture and also the Internet of Things. The objective of the thesis is to get an overview of the different data sources of the partners: what they contain and where they are stored.

The CITYZER projects develops new digital services and products to support decision making processes related to weather and air quality in cities. This includes, e.g., early warnings and forecasts, which allow for avoiding weather-related accidents, lessen human distress and costs from weather-related damage and poor air quality, and generally improve the resilience and safety of the society. The CityzerDEMO is a pilot demonstrating the Cityzer ecosystem in the Helsinki Metropolitan area. CityzerDEMO is currently being planned and implemented, and is expected to be opened for public in late 2017 or early 2018. (Cityzer official project plan 2015.)

The target groups of the services and products (e.g., public sector, real estate and energy companies, and distributors) and related business models will be analyzed and developed in collaboration with partner countries (India, Brazil and China) taking advantage of the pre-existing contacts by the FMI, Vaisala Ltd and CLIC Innovation. CITYZER project partners include Pegasor Ltd (air quality instrumentation and products), Vaisala Ltd (weather observation instrumentation and products), INNO-W Ltd (business services), Sasken Ltd (mobile products), Emtele Ltd (Portable IoT ICT Service Operation Center/Environment and remote intelligent cabinet for sensor network-GW and connections), HSY (urban services), Haaga-Helia University of Applied Sciences (service business models including digital services), Finnish Meteorological Institute (definition of and scientific research on meteorological & air quality products), and the Tampere University of Technology (defini-

tion of and scientific research on air quality products), as well as the CLIC Innovation Ltd as a subcontractor for information dissemination. (Cityzer official project plan 2015.)

The project is expected to have a wide variety of impacts and the pilot products and services presented by the project could be greatly beneficial. For example, the varying warnings could help to lessen weather-related damages and reduce air-quality related deaths, sick leaves and health care costs. The data provided by the project services could be used to guide rescue services and plan preventive actions. Likewise, the data provided could be used for designing more resilient, less sensitive and more secure solutions in city planning. Analysis, diagnostics and forecast tools need to be prepared to make useful predictions of future conditions, with access to the aforementioned data. (Cityzer official project plan 2015.)

2 The Research Question and Objectives

2.1 The Research Question

The research question is “What kind of data is available for analysing weather and air quality in the Helsinki region?

Sub questions relating to the research question are:

- Which data is provided by the Cityzer project partners and third parties?
- The weather and air quality data: where does it come from, what does it contain and where is it stored?
- How can the data be classified?

2.2 The Research Objectives

The different partners and companies will be interviewed to understand their role in the project. The main objective is to understand the weather and air quality data: where it originates from, what it contains and in which form and where it is stored. The final objective was to map and understand the data and the business ecosystem around it, and then classify the data and paint a picture of the whole ecosystem around the data.

3 The Theoretical Background

3.1 Weather and Air Quality

This chapter is about the different factors affecting weather and air quality.

3.1.1 Factors Affecting Weather

There are many factors that influence weather, many of which we cannot see, such as radiation, turbulence and air pressure. One of them is evaporation, the process of water rising in the form of water vapor from lakes, rivers, oceans, plants, the ground and other sources, at other times when the sun warms the surface of the Earth. Water vapor provides the moisture that forms clouds, eventually returning to Earth in the form of precipitation, and the cycle continues. The second factor is air masses; when air hovers for a while over a surface area with uniform humidity and temperature, it takes on the characteristics of the area below. These massive volumes of air often cover thousands of miles and reach to the stratosphere. Mid-latitude cyclonic storms and global wind patterns move them to locations far from their source regions. When two air masses meet, the cooler air pushes the warmer air upwards. When going up the temperature drops and the air can't hold as much water, as when it is warm, so the water molecules condense and form clouds. (Climate and Weather 2014.)

A jet stream is another factor and is it is a current of high-speed winds, and is usually found around five to ten miles above the Earth’s surface. The greater the temperature difference between the air masses, the greater the air pressure difference, and the faster the wind blows in the jet stream. This river of air has wind speeds which often exceed 100 mph, and sometimes over 200 mph. Jet streams more commonly form in the winter, when there is a greater difference between the temperature of the cold continental air masses and warm oceanic air masses. (Climate and Weather 2014.)

A weather front is the transition zone between two air masses of different humidity and temperature. Along a cold front, cold air displaces warm air and along a warm front, warm air displaces cold air. When neither air mass displaces the other, a stationary front develops. Towering clouds and intense storms may form along cold fronts, while widespread clouds and rain, snow, sleet, or drizzle may accompany warm fronts. (Climate and Weather 2014.)

3.1.2 Factors Affecting Air Quality

Pollution, more specifically, the amount and kind of pollutants released into the air, affects the air quality in a significant way. The factors affecting air quality include topography (terrain), such as mountains and valleys, weather (such as wind, temperature, air turbulence, air pressure, rainfall and cloud cover) and the physical and chemical properties of pollutants. Poor air quality can result from a combination of factors. Regional air quality is affected by how air behaves as a result of the interaction of topography and weather, and by the emission sources themselves. Air pollutants mix and disperse quickly in a large airshed because the air flow is not limited by topography but can travel and mix over great distances, resulting in good air quality in the air shed. Sometimes topography and weather combine to prevent pollutants from mixing and dispersing. In this case the pollutants become trapped within the area, resulting in poor air quality in the air shed. (BC Air Quality.)

Once pollutants are emitted into the air, the weather determines how well they disperse. Turbulence mixes pollutants into the surrounding air. For example, during a hot summer day, the air near the surface can be much warmer than the air above. Sometimes large volumes of this warm air will rise to great heights, resulting in vigorous mixing. Wind speed also contributes to how quickly pollutants are carried away from their original source, but strong winds don't necessarily disperse the pollutants, such as the smoke from open burning or forest fires. (BC Air Quality.)

Sometimes the condition of the atmosphere is very stable and there is very little vertical mixing. This occurs when the air near the surface of the earth is cooler than the air above, also known as a temperature inversion. This cooler air is heavier and doesn’t mix well with the warmer air above. Any pollutants released near the surface gets trapped and build up in the cooler layer of air near the surface. Temperature inversions are very common in mountain valleys, often forming during calm clear nights with light winds. They can even persist throughout the day during the winter. (BC Air Quality.)

Pollutants are released into the air from natural and human sources, from point and non-point sources. Many pollutants undergo chemical reactions when they encounter other pollutants in the air. The products of these chemical reactions are called secondary pollutants, as opposed to primary pollutants that are emitted directly into the atmosphere. Ground-level ozone is an example of a secondary pollutant that forms when nitrogen dioxide (NO^2) and volatile organic compounds (VOCs) mix in the presence of sunlight. (BC Air Quality.)

3.2 Collecting Weather Data

Weather data are collected by using different methods, such as data collection sheet, recording maximum temperature, rainfall and cloud data. ICT helps with data collection, data storage, data manipulation and data sharing. Below are examples of easily available and affordable digital recording tools, which aid in collecting weather data without the presence of meteorologist. The internet gives access to satellite images which shows all kinds of weather-related data. (Weather for schools.)

The Vaisala Automatic Weather Station AWS310 is a device designed for reliable, accurate environmental measurements. As a stand-alone weather data collection system, Vaisala AWS310 requires only a minimal amount of maintenance. With optional Vaisala Observation Network Manager NM10 software AWS310 users are able to remotely monitor and control the observation stations. AWS310 can also be customized to operate as part of the user's existing data collection system or AWS network. The functions are synoptic meteorology and climatological research, hydrology and urban meteorology. The AWS310 includes built-in algorithms that test each measurement to ensure quality. The minimum and maximum readings of every parameter are thoroughly tested, as are the step limits. The resulting logged meteorological data is saved on the external compact flash card, but can also be transmitted to a remote workstation as a real-time feed. (Vaisala 2014.)

Images 1 and 2. Vaisala's weather instruments. Fairuz Bhuiyan. 2016

A digital rainfall monitoring unit collects rain data digitally, using a tipping bucket. When the first bucket has a set amount of rain in it, it tips up and empties itself. The other bucket is then filled and the process repeated. The receiver can be placed inside where it can be easily read. The outside unit transmits a signal to the display unit inside. A digital temperature sensor is a remote monitor, which records temperature readings and collects data from several locations. It sends a radio signal with the readings from the sensor to the display unit. The device can record maximum, minimum and current temperatures and some models can also record humidity. After taking readings the unit is reset to start recording the new maximum and minimum. A digital wired thermometer is useful for recording two sets of data. One is recorded by a sensor in the unit, the other via a sensor on the end of a cable. These different readings are shown one above the other on the display. These readings can be the current, maximum and minimum temperatures. Different thermometer models reset these readings in different ways. They are reset after the temperatures had been recorded on a paper sheet. (Weather for schools.)

Spreadsheet programs are efficient tools, which examine and display the contents of the weather data. The programs find a maximum and minimum and create a variety of graphs using the data. The graphs can be used to look at the data to see if the numbers make sense. If there are any inaccurate numbers they appear in the graph. (Weather for schools.)

3.2.1 Measuring Air Quality

There are numerous methods for measuring air quality, including permanent monitoring stations in communities, mobile instrumentation (e.g. on a truck or airplane), and industrial stack monitoring. These monitoring stations measure the presence of contaminants in the air, such as carbon monoxide (CO), nitrogen dioxide (NO_2), ozone (O_3), particulate matter (PM2.5 and PM10), sulphur dioxide (SO_2), and hydrogen sulphide (H_2S). Contaminants are measured in one of two ways: either through continuous (real-time) or non-continuous (discrete) monitoring. (BC Air Quality.)

In continuous monitoring, air is constantly measured and the data is automatically transmitted to a central database. In non-continuous monitoring, contaminants collect on a filter or canister over a specified period of time, mostly up to six days. Then a technician collects the filter or canister and sends it to a certified laboratory for measurement and analysis. Continuous and non-continuous data are housed in a central data warehouse, where they are screened by data validation technicians for errors. Once the data have been validated, the data can then be used for reporting out to the public. (BC Air Quality.)

3.3 Helsinki's Situation

This chapter looks into the weather and air quality situation of Helsinki: the problems and how they are measured.

3.3.1 Air Quality Problems in Helsinki

Air quality in Helsinki is one of the best in Europe. However, recently there has been a concerning increase of air pollution to hazardous levels in the city. This is mainly caused by vehicular traffic, as its emissions are released at the street level and directly to air. Transport affects the air quality in Helsinki the most due to the hazardous fine particles and nitrogen oxide. In springtime, air quality gets worse due to road and street dust generated by traffic when road surfaces dry out. Exhaust gases are poorly diluted especially on winter days with no or little wind, and air quality may worsen in busy traffic and transport environments. (HSY 2011.)

Areas near busy roads and streets, as well as busy streets surrounded by tall buildings in the inner city, are the areas with the most pollution. Sensitive groups of people, including those suffering from asthma, elderly people suffering from the cardiovascular diseases or emphysema and children, may be adversely affected when air quality worsens. At times of poor air quality, they may alleviate symptoms by moving indoors and by avoiding strenuous outdoor physical activity. (HSY 2011.)

In spring and summer, ozone is generated in sunny weather and can spread from other areas to the Helsinki region in fairly high concentrations. Smoke spreads as a result of forest fires and the agriculture field burning beyond the national borders and it can occasionally worsen air quality in spring and summer. Wind and rain help to clean the air and, as a result, air quality is often best in autumn. (HSY 2011.)

Burning of wood occasionally worsens air quality in housing areas, as smoke produced by the burning of wood in fireplaces, results in high concentrations of particle pollution in the air during cold weather when there is no or little wind. The effect can be severe especially in densely built areas of detached and semi-detached housing. According to the environmental regulations, only dry and clean wood can be burned in fireplaces. It is forbidden to burn twigs and waste at properties. Appropriately stored, dry wood and good burning techniques significantly decreases emissions caused by burning wood. The Helsinki Region Environmental Services Authority has published a guidebook, which gives instruc-

tions on wood burning, guides house owners and tenants in the Helsinki Metropolitan Area on the proper use and maintenance of fireplaces. (HSY 2011.)

3.3.2 How the Air Quality in HSY is Measured

Helsinki Region Environmental Services Authority (HSY) is a federation of municipalities in the Helsinki metropolitan area (cities of Helsinki, Espoo, Vantaa and Kauniainen), specializing in waste and water management as well as regional and environmental information. In the field of regional and environmental information, mitigation and adaptation of the climate change is one of their duties. HSY promotes and monitors the implementation of the Climate Strategy 2030 for the Helsinki metropolitan area, calculates the greenhouse gas emissions of the area annually, promotes the implementation of the Climate Adaptation Strategy in cooperation with cities and other regional operators and promotes material efficiency by developing the tools for waste monitoring. Another responsibility is the providing of the regional data, which involves the GIS data, e.g. information about housing, work places and land reserves. The third responsibility is air quality: monitoring, informing the public on air quality, air quality research and planning as well as air quality communication and education. HSY also takes care of outdoor air quality monitoring on behalf of the energy companies of Helsinki metropolitan area, the Port of Helsinki and Finavia at the Airport. (Niemi 25 April 2016.)

In the Helsinki metropolitan area, air quality has improved clearly (especially sulphur dioxide SO_2, carbon monoxide CO, nitrogen oxides NOx and lead Pb) during last 10-20 years. However, there are still problems with several pollutants, especially thoracic (PM10) and fine (PM2.5) particles as well as gaseous nitrogen dioxide (NO_2) and ozone (O_3).
The concentrations of nitrogen dioxide (NO_2) are high in areas with busy traffic, mainly due to emissions from diesel vehicles. The annual limit value of nitrogen dioxide is exceeded along the busiest streets, especially in street canyons. The hourly limit value exceedance is unlikely but possible. The city of Helsinki is now (year 2016) preparing a new air quality action plant to reduce NO_2 concentrations as well as PM10 and PM2.5 concentrations, focusing on mitigation of emissions from traffic and residential wood burning. The action plant for period 2008-2016 was not efficient enough to reduce NO_2 concentrations below annual limit value. (Niemi 25 April 2016.)

Table 1. Air pollutants and emission sources. HSY. 2015

Air pollutants	Emission sources in the area	Condition
Fine particles	traffic, long-range transport, wood burning, street dust	☹
Thoracic particles	street dust	☹
Ozone	long-range transport	☹
Nitrogen dioxide	traffic	☹
Benzo(a)pyrene	wood burning	😐
Carbon monoxide	traffic	☺
Sulphur dioxide	energy production, ships	☺
Heavy metals	traffic, industry	☺

The concentrations of thoracic particles (PM10), are high due to road dust, especially in spring. The use of studded winter tyres as well winter-sanding of streets causes formation and accumulation of dust material in winter. In spring streets dry out and subsequently large amounts of PM10 dust is released into the air. The EU's 24 h limit value for PM10 has been exceeded along the busiest streets of Helsinki, but not after 2006. The concentrations of thoracic particles in the air have fallen below the limit values of EU during the recent years, which has mainly been achieved by intensive dust binding and street cleaning measures (more information: www.redust.fi). (Niemi 25 April 2016.)

Fine particles (PM2.5) are the air pollutant that is the most detrimental to people. At a national level, their concentration is low. However, elevated concentrations are observed near busy streets and in densely constructed single-family housing areas due to wood burning. The World Health Organization (WHO) guideline values of fine particles, PM2,5, are exceeded due to long-range transport, local traffic, wood burning and fireworks. The EU's target value of benzo(a)pyrene is also exceeded in some small house areas due to wood burning. There are large spatial and temporal variation in fine particle concentrations, sources, compositions and size distributions, which poses a challenge for health impact assessments. (HSY 2015.)

The amount of ozone (O_3) increases in the spring and summer, especially outside the urban areas. Ozone is long-range transported to Finland from elsewhere in Europe with the winds. There are fairly high average concentrations of ozone and long-term objectives are, thus, exceeded. The concentrations of heavy metals are low. They fall below the target and limit values. The concentrations of some air pollutants have decreased significant-

ly and they no longer cause notable air quality problems in the Helsinki metropolitan area. Such pollutants are sulphur dioxide, lead and carbon monoxide. (HSY 2015.)

Table 2. Measured components in 2016. HSY. 2016

Measurement site	Representativeness	PM_{10}	$PM_{2,5}$	NO_x	SO_2	O_3	BC	PNC	VOC	PAH
Mannerheimintie	traffic site in Helsinki city centre	x	x	x			x			
Mäkelänkatu	street canyon in Helsinki	x	x	x		x	x	x	x	x
Kallio	urban background in Helsinki	x	x	x	x	x	x	x	x	x
Vartiokylä	small house area (wood burning)		x	x		x				x
Leppävaara	traffic site in Espoo	x	x	x						
Tikkurila	traffic site in Vantaa	x	x	x			x		x	
Luukki	rural background		x	x	x	x	x			
Vuosaaren satama	harbour		x	x	x					
Puistola	small house area (wood burning)		x	x						x
Lintuvaara	small house area (wood burning)		x	x						x
Hämeenlinnanväylä	main road	x	x	x						
Hernesaari	harbour				x					

Red = Measurement sites change every calendar year

Time resolution 1 h (average of minute values)

Time resolution 1 month

The current air quality monitoring and research priorities are the size distributions, compositions and sources of particles, the effectiveness of the pollution reduction measures, concentration trends, exposure and health hazards, new monitoring methods, mobile measurements as well as condensed sensor network and fusion modelling of pollution concentrations, including air quality forecast. In practice, the research and development work is conducted as co-operation projects, together with several research organizations and companies. For instance, in the CITYZER project the sensors of Pegasor are tested at HSY's measurement stations in Lintuvaara small house area in Espoo and Mäkelänkatu street canyon (super site measuring station) in Helsinki. FMI is also testing Pegasor sensor and it is stationed at an urban background site in the Kumpula area (III SMEAR measuring station) in Helsinki. Passive sampling methods for NO_2, SO_2 and volatile organic compounds (VOC) are indicative methods, used for surveys and mapping, weekly, monthly, annual averages. They are cheap, easy to use and analyzed in a laboratory.
The air quality impacts of local and regional emission sources are also assessed using dispersion models. HSY uses OSPM model for modelling air quality in street canyons. FMI has developed models for open road environments (CAR-FMI), the whole city level (UDM-FMI) as well as for regional and long-range transport and forecasts (SILAM). For instance, online SILAM forecasts help to anticipate and evaluate the long-range transport episode situations, including pollutants from ordinary anthropogenic sources and forest fires. (Niemi 25 April 2016.)

The impacts of air pollutants can also be estimated by bioindicators. There are over 700 areas in the Uusimaa region, which are measured every 5th-10th year. The biodindicator assessments are financed by the companies and municipalities and the damages can be seen from the lichens (*epiphytic lichens*). (Niemi 25 April 2016.)

HSY publishes notifications on air quality every weekday, and takes action upon sudden deteriorations in air quality. A special notice is published when the concentrations rise higher than normal, for instance if the PM10 daily limit value 50 µg/m^3 is exceeded or if air quality is poor several hours according to air quality index. There are five air quality categories ranging from good to very poor. The HSY index is based on limit and threshold values and guidelines, and therefore reflects health impacts of air pollutants. Health problems can arise in sensitive individuals when air quality is poor. Air quality is measured continuously and the results are real-time (hourly-daily) in the web. The information is uploaded on information channels on weekdays, the press, radio, TV, Internet and information screens in trams, metros and in selected railway stations. (Niemi 25 April 2016.)

3.4 Big Data

Because the Cityzer project has features of the Big Data phenomenon, the Big Data Framework is selected for the theoretical framework of the thesis.

3.4.1 Introduction to Big Data

The term ’Big Data’ is often used to describe data sets whose size is beyond the capability of commonly used software tools to capture, manage and process data within a tolerable period of time. The Big Data paradigm consists of the distribution of data systems across horizontally coupled, independent resources to achieve the scalability needed for the efficient processing of extensive datasets. The data life cycle is the set of processes in an application that transform raw data into actionable knowledge. Big Data is being generated by everything surrounding it. It is produced by every digital process and social media exchange and transmitted by systems, sensors and mobile devices. (Pritzker & May 2015a, 4.)

Big data has three or four characteristics, known colloquially as the “four V’s”. The first one is volume, which defines the large amount of both open and private data. The second characteristic is velocity, which means the fast rate that data is received and analysed. Variety stands for the high variety in data sources and formats, such as new unstructured and semi-structured data types, such as text, audio, image, video and sensor data. The

last characteristic, value, refers to the extracted, intrinsic value of the data. The Big Data paradigm consists of the distribution of data systems across horizontally coupled, independent resources to achieve the scalability needed for the efficient processing of extensive datasets. (Pritzker & May 2015a, 4.)

Big data is creating a new culture in which business and IT leaders must join forces to realize value from all data, thus changing the way people within organizations work together. With the help of big data, employees are able to make better decisions, engage customers, optimize operations, prevent threats and fraud, and capitalize on new sources of revenue. Data is emerging as the world's newest resource for competitive advantage. Because of big data and the Internet of Things (IoT), which connects physical and digital worlds, new technological innovations have emerged. These include the maximum efficiency for heating and air conditioning using presence detection, weather predictions and remote control, intelligent road services, transportation on the demand, route and schedule optimization, mobile shopping and payments, new security recognition systems, smart deliveries (tracking and monitoring of parcels) and monitoring patients at home, among other smart innovations. (Roponen 2016.)

The benefits of connected people and devices will change over time, as connectivity becomes a commodity, and through higher experience & demand of the real value. These include improved, real-time customer interaction, revenue from new application and services, reductions in warranty related costs due to remote diagnostics and differentiation opportunities from content and service perspective. (Roponen 2016.)

3.4.2 Mapping Big Data

Big data is an approach to working with data, the Location Analytics tools required to work with it, and derive business value. By bringing together big data and mapping, businesses and organizations can get a lot of benefits. They can drive faster time to market by highlighting previously unseen patterns and relationships within existing data sets. They can also bring together different technologies like Business Intelligence (BI) and Customer Relationship Management (CRM) with Location Analytics to deliver enhanced business insight. The combination of big data and mapping can ultimately lead to more accurate decision-making as well as act as a driver of business strategy, potentially delivering enhanced customer engagement, improved profitability and greater competitive advantage. (Green 2014.)

3.4.3 Metadata

Metadata is described as “data about data” and it is an important concept to Big Data. Metadata describes additional information about the data such as how and when data was collected and how it has been processed. Metadata has all the requirements for tracking, change management, and security. Provenance is the history of a dataset and provenance type of metadata guides users to correct data utilization when the data is repurposed from its original collection process in an effort to extract additional value. Semantic metadata, another type of metadata, refers to the description of a data element to assist with proper interpretation. (Pritzker & May 2015a, 12.)

An ontology is a graphic model, representing a semantic relationship between entities. Ontologies are semantic models constrained to follow different levels of logic models. Ontologies can be very general or extremely domain-specific in nature. Taxonomies are metadata about data element relationships. Taxonomy is a hierarchical relationship between entities, where a data element is broken down into smaller component parts. (Pritzker & May 2015a, 12.)

3.4.4 Big Data Analytics

Big Data analytics is the ability to process large amounts and various types of information. Big data analytics examines large amounts of data to uncover hidden patterns, correlations and other insights. It enables the analyzation of data and getting answers from it directly, compared to more traditional business intelligence solutions. In-memory analytics is a technology, where data is analysed from system memory, enabling the removal of data preparation and analytical processing latencies to test new scenarios, create models and run iterative and interactive analytics scenarios. (Pritzker & May 2015a, 4.)

Big data analytics helps organizations to support their data and use it to identify new opportunities, resulting in smarter business moves, more efficient operations, higher profits and happier customers. There are a lot of benefits of big data analytics. One of them is cost reduction by using big data technologies and cloud-based analytics, because they can store large amounts of data and identify efficient business opportunities. The second is faster and better decision making due to the speed of in-memory analytics, combined with the ability to analyse new sources of data right away. The third is providing customers with new products and services and meeting their needs. (SAS Institute Incorporated.)

3.4.5 The Big Data Architecture

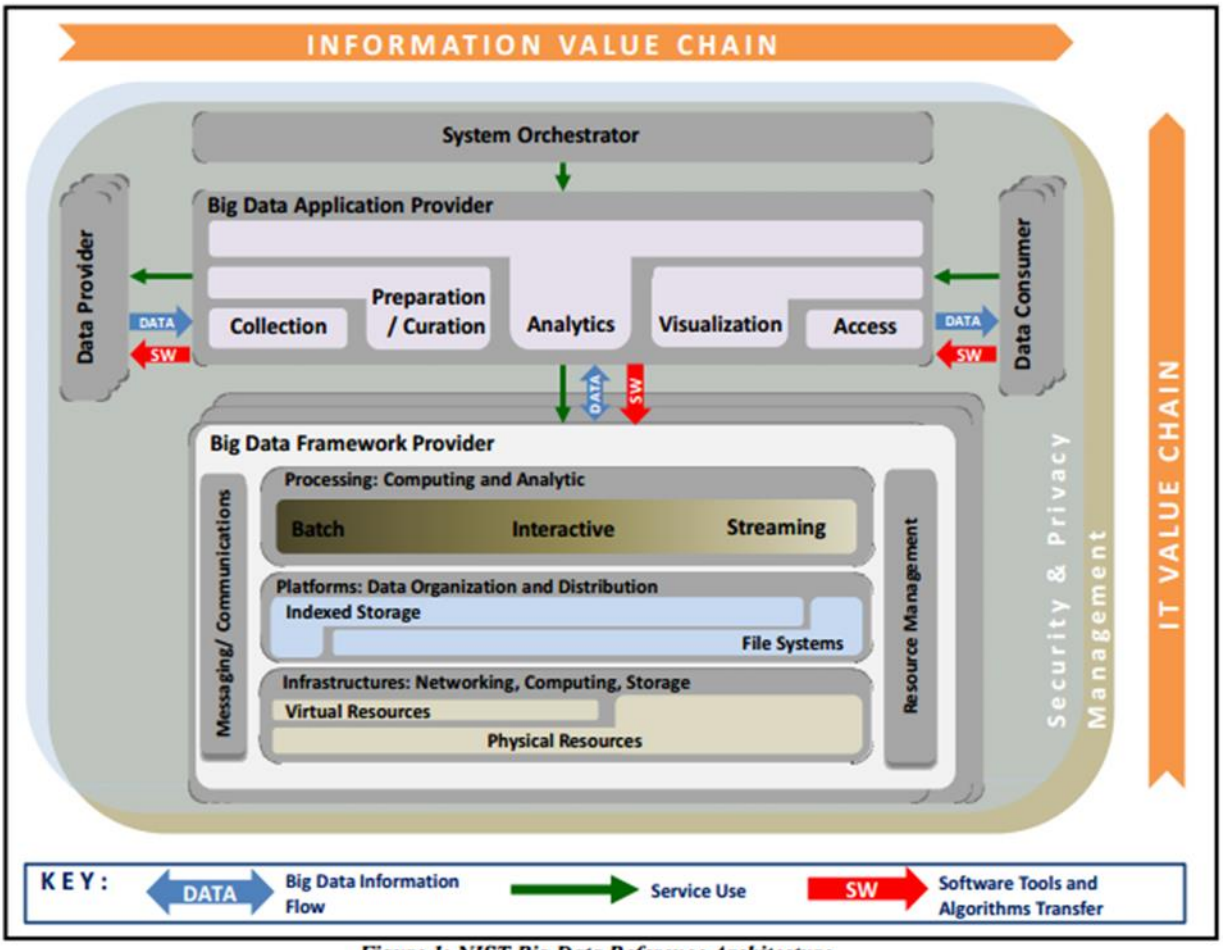

Figure 1. NIST Big Data Reference Achitecture. Pritzer & May. 2015

In system architecture, actors and roles have a significant relationship, just like in the silver screen. The roles are the parts played by multiple actors. System development actors represent individuals, organizations, software, or hardware. Examples of actors can be sensors, applications, software agents, individuals, organizations, hardware resources and service abstractions. These seven items, five main architecture components and two fabrics interwoven in them, form the foundation of the reference architecture taxonomy. The five main components, which represent the central roles, are System Orchestrator, Data provider, Big Data Application Provider, Big Data Framework Provider and Data Consumer. (Pritzker & May 2015b, 11.)

A System Orchestrator defines and integrates the required data application activities into an operational vertical system. The System Orchestrator provides the overarching requirements that the system must fulfil, including policy, governance, architecture, re-

sources, business requirements and monitoring or auditing activities to ensure that the system complies with those requirements. The System Orchestrator role provides system requirements, high-level design, and monitoring for the data system. The actors are business leadership, consultants, data scientists and information, software, security, privacy and network architects. The activities are business ownership, governance, data science and system architecture requirements and monitoring. (Pritzker & May 2015b, 12-13.)

A Data Provider introduces new data or information feeds into the Big Data system. The Data Provider makes data available to itself or to others. The actor fulfilling this role can be part of the Big Data system, from another system, or internal or external to the organization orchestrating the system. Once the data is within the local system, requests to retrieve the needed data is made by the Big Data Application Provider and routed to the Big Data Framework Provider. The Data Provider captures data from its own sources or others, in other words, it captures data from a data producer, which may be a sensor or an organizational process. Aspects of the data sources activity include both online and offline sources. Online sources may be web browsers, sensors, deep packet inspection devices or mobile devices and offline sources are possible public records or internal records. The actors are usually enterprises, public agents, researchers and scientists, search engines, WEB, FTP and other applications, network operators and end users. The activities may be data collection from sources, data persistence, data scrubbing and data annotation, access rights management and access rights policy contracts, data distribution APIs and data availability publication. (Pritzker & May 2015b, 15-16.)

A Big Data Application Provider executes a life cycle to meet security and privacy requirements as well as System Orchestrator-defined requirements. This is where the general capabilities within the Big Data framework are combined to produce the specific data system. The actors are application specialists, platform specialists and consultants. The activities are collection, preparation, analytics, visualization and access. (Pritzker & May 2015b, 18.)

A Big Data Framework Provider established a computing framework, in which to execute certain transformation applications, while protecting the privacy and integrity of data. It has general resources or services to be used by the Big Data Application Provider in the creation of the specific application. The Big Data Application Provider chooses the technologies and uses these resources and network to build the specific system. The actors are in-house clusters, data centres and cloud provider. These activities are infrastructure, data platforms and processing frameworks. (Pritzker & May 2015b, 21.)

A Data Consumer includes end users or other systems who use the results of the Big Data Application Provider. It receives the value output of the Big Data system. After the system adds value to the original data sources, the Big Data Application Provider offers that same functionality to the Data consumer. The actors are end users, researchers, applications, and systems. These activities are searching and retrieving, downloading, analysing logically, reporting and visualizing. (Pritzker & May 2015b, 22-23.)

The Management Fabric of Big Data systems handle both system- and data-related aspects of the Big Data environment. It encompasses two general groups of activities, which are system management and Big Data life cycle management. System management includes activities such as provisioning, configuration, package, software, backup, capability, resources, data and performance management. Big Data lifecycle management involves activities surrounding the data life cycle of collection, preparation/curation, analytics, visualization and access. The actors are in-house staff, data centre management and cloud providers. (Pritzker & May 2015a, 23-24.)

The Security and Privacy Fabric interacts with the System Orchestrator for policy, requirements and auditing, and also with the Big Data Application Provider and the Big Data Framework Provider for development, deployment, and operation. The activities are security and privacy policy requirements and security and privacy monitoring. The actors are corporate and security officer and security specialist. (Pritzker & May 2015a, 24.)

3.5 Data, Information, Knowledge and Wisdom

Computers are often called data processing machines or information processing machines, designed for the input, storage, processing, and output of data and information. They are also knowledge processing machines and have wisdom. Data is information, often in the form of facts or figures, such as numbers, text, images and sounds, obtained from experiments or surveys, used as a basis for making calculations or drawing conclusions. Data is stored and processed by a computer. Information is the definite knowledge acquired or supplied about something or somebody. It refers to the collected facts and data about a particular subject and the communication of facts and knowledge. Knowledge is the general awareness or possession of information, facts, ideas, truths, or principles learned throughout time. It is familiarity or understanding gained through experience or study. Wisdom is the knowledge and experience needed to make sensible decisions and judgments, or the good sense shown by the decisions and judgments made. It is the gathered knowledge of life or in a particular sphere of activity that has been gained through experience. (OTEC.)

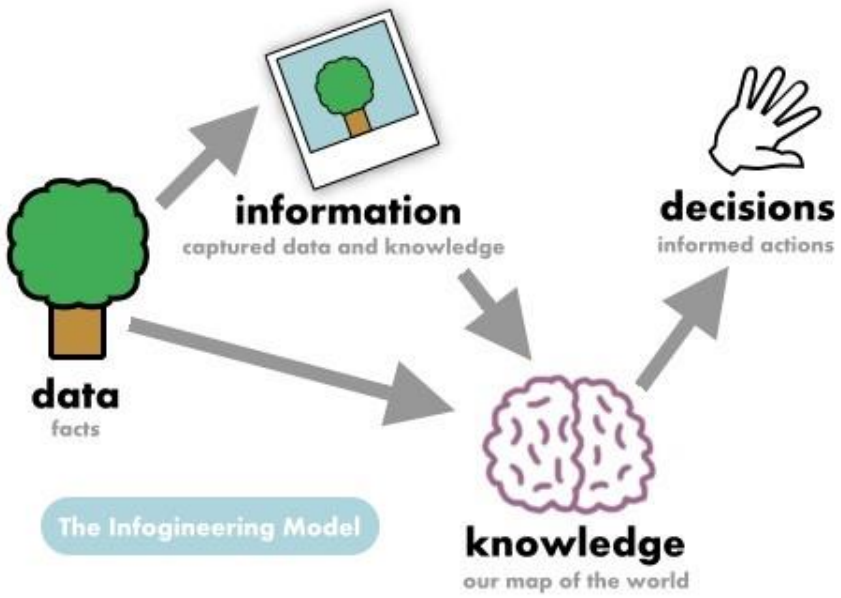

Figure 2. Data, Information, Knowledge, Decisions model. Ian Dunbar's Weblog. 2013

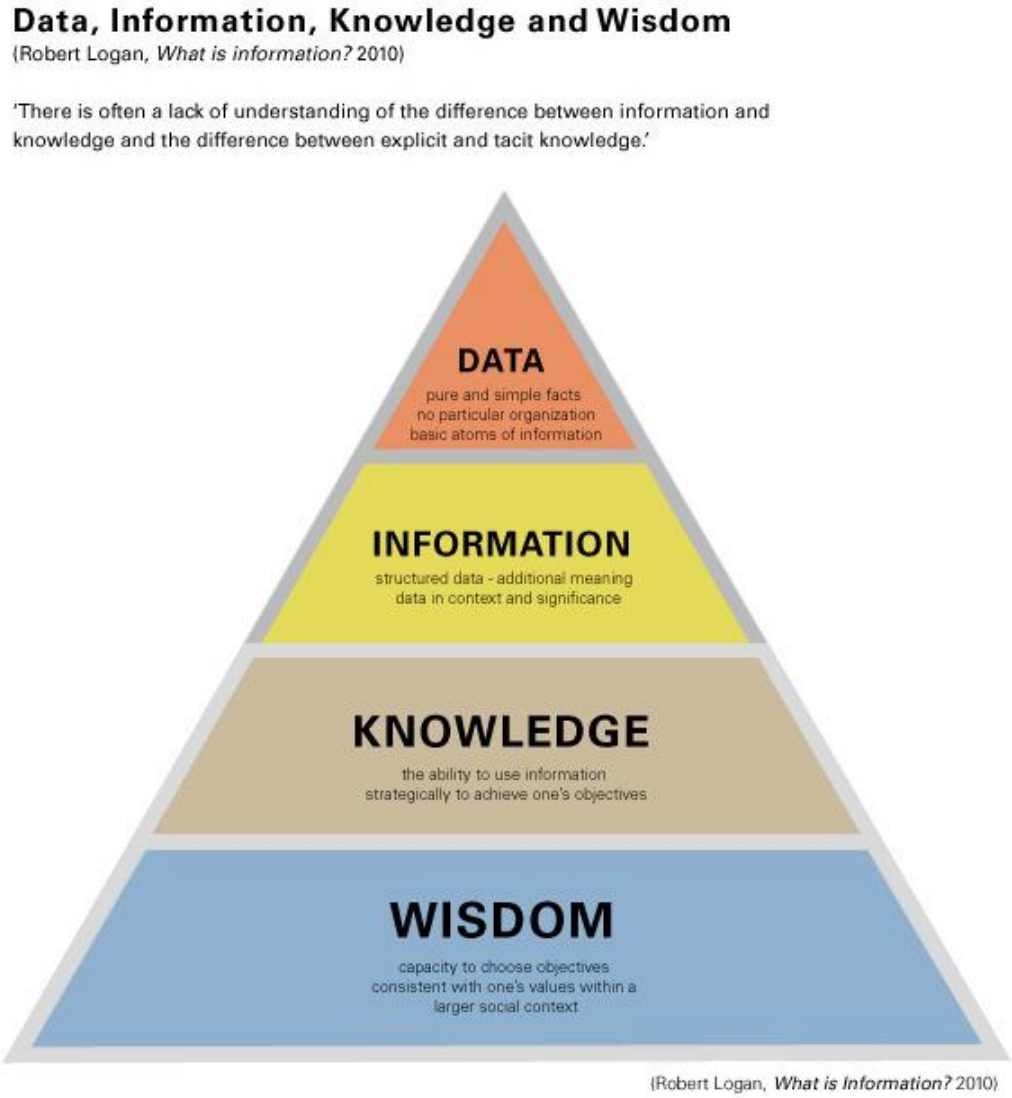

Figure 3. What is Information?. Logan 2010

3.6 The Internet of Things (IoT)

The Cityzer project also has features of the IoT, and so the IoT framework is also included in the thesis.

3.6.1 The Definition of IoT

The Internet of Things (IoT) is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction. A thing, in the Internet of Things, can be a person with a heart monitor implant, a farm animal with a biochip transponder, an automobile that has built-in sensors to alert the driver when tire pressure is low or any other natural or man-made object that can be assigned an IP address and provided with the ability to transfer data over a network. Practical applications of IoT technology can be found in many industries today, including precision agriculture, building management, healthcare, energy and transportation. An increase in the number of smart nodes, as well as the amount of upstream data the nodes generate, is expected to raise new concerns about data privacy, data sovereignty and security. (IoT Agenda 2014.)

IoT has its roots from the convergence of wireless technologies, micro-electromechanical systems (MEMS), microservices and the Internet. The convergence has helped break down the barrier between operational technology (OT) and information technology (IT), allowing unstructured machine-generated data to be analyzed for insights, resulting in improvements. The concept has been in development for decades, but named in 1999. One of the first IoT objects were ATM's, invented in 1974. The first Internet appliance was a Coke machine at Carnegie Melon University in the early 1980s. The programmers could connect to the machine over the Internet, check the status of the machine and determine whether or not there would be a cold drink awaiting them, should they decide to make the trip down to the machine. (IoT Agenda 2014.)

According to statistics, in 2008, there were more objects connected to the internet than people. By 2020, 250 000 vehicles will be connected to the Internet. The IoT is rumoured to add 8-13 trillion euros to global GDP in the next 20 years. (Morgan 2014.)

3.6.2 The Internet of Things and Weather Forecasting

The Internet of Things will help with the prediction of weather condition and attaining higher accuracy and flexibility. Remote sensing technology has enabled the real-time analysis

of weather data and has transformed the way that was used to collect and analyse weather data and build a strong database for reliable weather forecasts. IoT enabled monitotring systems are designed to collect data from various vehicles moving on the road, as they will transmit the weather and road condition data (temperature, pressure, moisture and light sensors as well as motion sensors such as accelerometers and gyroscopes), which would help build more accurate forecast and provide flexible real time monitoring in different time zones. (Mahendra 2016.)

Sensors are installed on windshields, wipers and tyres of cars. These sensors in integration with IoT help in collecting weather data. Companies like IBM, Rainmachine and others are working towards the expansion of IoT enabled weather forecasting. Weather forecasting is the application of science and technology to predict the state of the atmosphere for a given location. (Mahendra 2016.)

The accuracy of weather forecasting directly or indirectly influences other sectors of economy to a great extent, raising the need of a system that facilitates higher accuracy of real time monitoring and future weather prediction. The IoT weather forecasting technology can benefit agriculture, as the agricultural process, such as the preparation of soil, sowing, irrigation, harvesting and storage of crops, are directly dependent on weather conditions. This leaves farmers vulnerable to weather hazards, but with the IoT technology, vital weather prediction can be delivered to farmers, and they can then use the intelligence to improve their crop fertility and cost along with taking essential steps to diversify weather hazards. Timely and accurate delivery of weather forecast will ensure higher productivity and lower the risk of weather hazards. (Mahendra 2016.)

Transportation would also greatly benefit from IoT, as the successful installation of remote sensors on every vehicle in motion, would immediately transmit information about every minor detail of a change in weather for analysis, allowing the real time weather monitoring and forecasting report to cover details like temperature, fog, road condition, light, flood, storm and other conditions that would add up to reliability and accuracy of the report. It would promote smart and intelligent driving systems and improve the safety and security of drivers, resulting in a decrease in road accidents. (Mahendra 2016.)

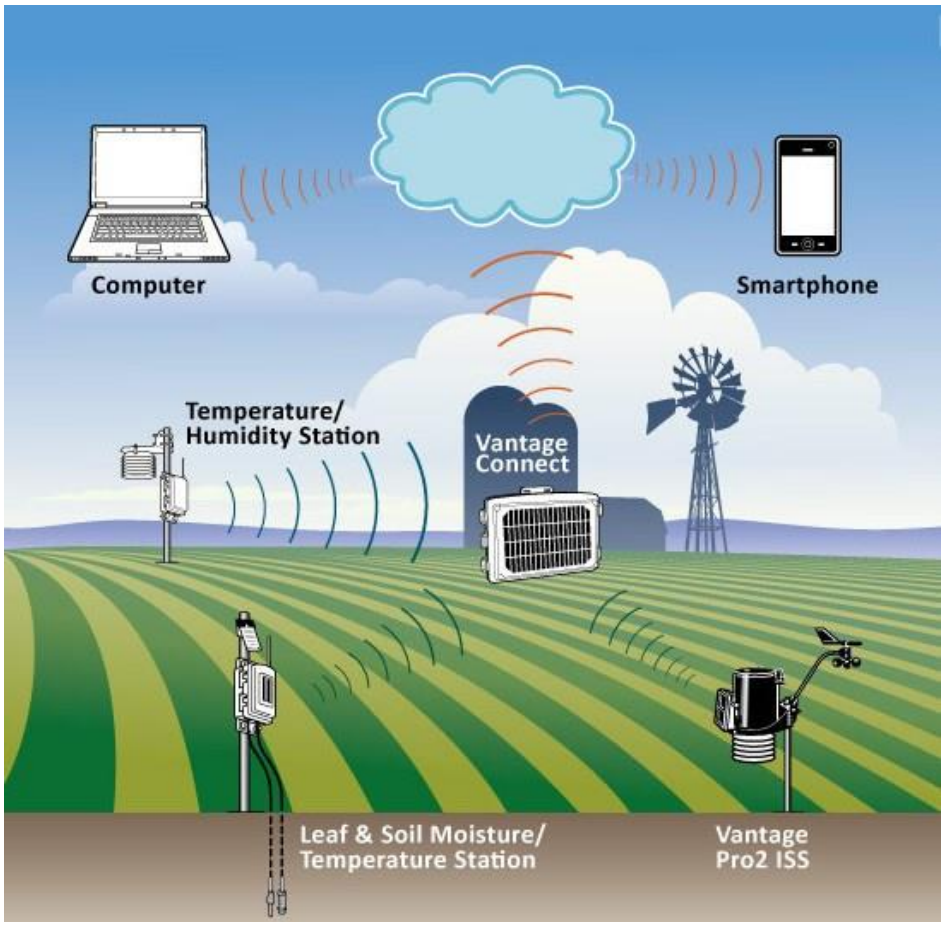

Figure 4. An illustration of a remote weather station. IoT Worm. 2016

4 The Research Methodology

4.1.1 Data

This thesis is based primarily on secondary data but also on primary data, which includes qualitative and quantitative research. Secondary data are data that have been interpreted and recorded, in the form of magazines, newspapers and articles on the Internet. This thesis relies on Internet sources a lot, because it contains authorative articles and useful information. The information is gathered from the internet, such as from thesis reports, scientific articles and also from literature related to the topic.

Primary data are the most recent recording of a topic/situation and can provide personal and reliable information. It is a combination of qualitative and quantitative data. Quantitative data are concrete data, which refers to numbers that can be analysed in the form of statistics and be measured quite accurately. Qualitative data are abstract data that can't be measured, and are expressed in words instead of numbers. This type of data is based on people's opinions and ideas. (Walliman 2011, 70-73.)

4.2 Interview

A semi-structured interview will be conducted in a focal group consisting of experts from the partners Vaisala, Finnish Meteorological Institute, Helsinki Region Environmental Services Authority (HSY), Pegasor and Tampere Univeristy of Technology (TUT) to gain valuable information required for the thesis project. Conducting interviews provides the researcher with adequate information and valuable qualitative data. The use of interviews to question samples of people is a very flexible tool with a range of applications. Interviews can be used for subjects, both general or specific in nature and even, with the other correct preparation, both general and specific in nature for sensitive topics. The interviewer is in a good position to judge the quality of responses, to notice if a question has not been properly understood and to encourage the respondent to give complete answers. (Walliman 2011, 99.)

There are three types of interviews: structures, unstructured and semi-structured interviews. Structured interviews concentrate on standardized questions read out by the interviewer according to an interview schedule. Answers may be closed format. An unstructured has a flexible format, usually based on a question guide but where the format remains the choice of the interviewer, who can allow the interview to express themselves freely and ramble on for hours in order to get insights into the attitudes of the interviewee.

The questions may not be closed format. A semi-structured interview contains both structured and unstructured questions with standardized and open type questions. (Walliman 2011, 97-99)

5 Results

5.1 The Leading Partners of the Cityzer Project

The organizations and companies that were interviewed were the 4 main partners of the project: The Finnish Meteorological Institute (FMI), Helsinki Region Environmental Services Authority (HSY), Vaisala, Pegasor and Tampere University of Technology.

The Finnish Meteorological Institute (FMI) is a research and service agency under the Ministry of Transport and Communications. The main objective of the Finnish Meteorological Institute is to provide the Finnish nation with the best possible information about the atmosphere above and around Finland, for ensuring public safety relating to atmospheric and airborne hazards and for satisfying requirements for specialized meteorological products. (Finnish Meteorological Institute.)

The Helsinki Region Environmental Services Authority (HSY) is a municipal body, which produces waste management and water services, as well as providing information on the Helsinki Metropolitan Area and environment, helping inhabitants to act for a better environment. The member cities of the Helsinki Region Environmental Services Authority HSY are Espoo, Helsinki, Kauniainen and Vantaa. (HSY 2016.)

Vaisala is a Finnish company, located in the city of Vantaa, that develops, manufactures and markets products and services for environmental and industrial measurement. The major customer groups and markets are national meteorological and hydrological services, aviation authorities, defence forces, road authorities, the weather critical energy sector, life science and high-technology industries and building automation. Vaisala serves customers in over 140 countries. (Wikipedia 2016.)

Founded in 2008, Pegasor is a Tampere-based leading cleantech company in the field of particle emission monitoring. Pegasor provides its worldwide customers value adding features provided by the company’s particulate sensor technology. Pegasor’s sensors specialize in continuous real-time monitoring of fine and ultrafine particles, in applications like engine emissions, stack emissions, indoor air quality and ambient air quality. The sensors also monitor particle number and mass concentration. Pegasor products are widely in use in these industrial applications by leading companies and researchers. (Pegasor.)

Tampere University of Technology is one of the two Finnish universities operating in the form of a foundation, located in Tampere, in the suburb of Hervanta. The foundation mod-

el promotes the development of education and research at TUT and gives the University good prerequisites to succeed amid growing international competition. The proceeds of the 137-million-euro foundation capital enable further investment in new openings in research and education as well as in the development of quality and operations management. TUT's campus is a community of 10,500 undergraduate and postgraduate students and close to 2,000 employees. (Tampere University of Technology.)

5.2 Key Findings

5.2.1 Interview

Interviewing was chosen as the main method, as it was the most efficient way to find out questions to the answers from the organizations of the Cityzer project. The aim was to find out about the organizations' data sources and analyze the results. The questions asked were semi-structured questions, as the interviewees were also allowed to express their opinions freely and provide background information on the topic. The interviewing period lasted for a month and was done by visiting the headquarters of companies and organizations. Pegasor was interviewed via phone and Tampere University of Technology via e-mail, as both organizations were based in the city of Tampere. The format of the interview was standardized and open-ended. The same open-ended questions were asked to all interviewees, as it enabled faster interviews that were easily analyzed and compared. All of the interviews were recorded, with permission from the interviewees. Below is an overview of the key findings concluded from the interviews.

5.2.2 The Role of the Organizations/Companies in the Cityzer Project

The Finnish Meteorological Institute (FMI)

The Finnish Meteorological Institute is the main founder of the Cityzer-project and Ari Matti Harri is the project owner. The FMI has over 600 employees, out of which a small group is involved in the Cityzer project. The role of FMI in the data architecture is both the Data Provider and System orchestrator. (Nousiainen 20 April 2016.)

Helsinki Region Environmental Services Authority (HSY)

In general, HSY also supports the development of products and services that utilizes e.g. Emtele's wireless data transport applications, Vaisala's Observation Network Management (NM10), FMI's air quality fusion model and Sasken's mobile services. (Niemi 25 April 2016.)

Figure 5. Monitoring stations in Helsinki. HSY. 2016

HSY acts as a data source and end user, does research in co-operation with the other organizations to investigate properties and dispersion of pollutants from traffic and wood burning. It is responsible for air quality monitoring in the Helsinki metropolitan area and thus has a big role as the data source. It is a key partner for the CityzerDemo in Helsinki. Regarding the data architecture model, its main role is the data provider, but also the consumer of model results. HSY didn't apply for funding from TEKES, as it funded this project by itself, by doing some work for it and investing in it. HSY invested in this project because there were some important aspects regarding air quality in the Helsinki metropolitan area-capital region, which are researched, developed and demonstrated. (Niemi 25 April 2016.)

Pegasor

The role of Pegasor in the Cityzer project is a Data Provider. The company provides air quality information on particles and their behavior information to other partners and companies, so that they can benefit from it. (Saukko 2 May 2016.)

Vaisala

Vaisala is the device provider, so it produces weather quality measuring devices and sells them to other companies and organizations, such as the Finnish Meteorological Institute and it extract the data from the device. The only data the company provides are flash data collected from the flash tracking device. In the future, Vaisala is hoping to get the role of the System Orchestrator for the Cityzer project, so it would have the opportunity to see the

big picture in the whole “chain”. It could then create new services and bring value to the “chain”. (Laakso, Ståhle & Ojanperä 18 May 2016.)

Tampere University of Technology (TUT)

TUT’s role is, by its expertise and activities in traffic emissions and air quality research, to support the research on air quality sensor networks. This means e.g. conducting experiments and modeling on traffic emissions in urban environments, calibration and comparison studies for particle sensors and participation in workshops. It is a Data Provider for the Cityzer project. Topi Rönkkö, who was interviewed, corresponds to TUT’s research in the Cityzer project. (Rönkkö 7 June 2016.)

5.2.3 The Data Source

The Finnish Meteorological Institute (FMI)

Air quality is measured by the In situ method by physical contact. Smaller particles are more important. The data within the Cityzer system is divided into three classes: (1) observational data, (2) model data, and (3) metadata. These data have different sources and usages. The primary purpose of observational data is to serve as input for nowcast and forecast models. These models, essentially, produce “observations” for the future. (Nousiainen 20 April 2016.)

Two different kinds of nowcasts are produced. The first is a deterministic nowcast, provided by the system, which presents the most likely development of precipitation over the next three hours. The second is a probabilistic nowcast, which enables the estimation of probability that precipitation exceeds some certain threshold conditions. As an example, the probabilistic nowcast determines areas and times where precipitation exceeds 10 mm/h with 50% probability. In addition to predicting the amount (water-equivalent accumulation) of precipitation as a function of place and time, also the type of precipitation is predicted. (Nousiainen 20 April 2016.)

The SILAM model provides the so-called background air quality data for the Cityzer air quality forecast, including long-range transport of pollutants to the Cityzer domain from sources far away. The SILAM model has been developed at the FMI and is also run at the FMI. Finally, input model data are used to derive the precipitation type (snow, rain, hail). The meteorological data needed (e.g., temperature and humidity profiles within the Cityzer do-main) for this are also obtained from the HIRLAM NWP model. These data are used by

the RAVAKE nowcasting model to correctly predict the precipitation type in the precipitation nowcasts. (Nousiainen 20 April 2016.)

Helsinki Region Environmental Services Authority (HSY)

HSY is responsible for the air quality monitoring in the Helsinki metropolitan area. There are 11 air quality monitoring stations, out of which 7 are continuous and 4 change their locations every calendar year. HSY's air quality monitoring stations are located in different environments, including traffic sites, small house areas, harbors as well as urban and rural background sites. Online monitors are used to measure gaseous (e.g. nitrogen dioxide NO_2, nitrogen monoxide NO, ozone O_3, sulphur dioxide SO_2) and particulate pollutants (e.g. thoracic particle mass PM10, fine particle mass PM2.5, black carbon BC, particle number concentration). They record data with minute time resolution. The monitoring devices are connected to the main computer of each measurement station and the minute data it sent wirelessly to central computer located at HSY's office premises. HSY has a data collection system for data storing. HSY's air quality results of the Helsinki metropolitan area are shown with 1-h time resolution at the web pages of HSY (www.hsy.fi/en/residents/theairyoubreathe). (Niemi 25 April 2016.)

Most of the data is also sent to FMI and that data is presented in the Finnish air quality portal (http://www.ilmanlaatu.fi/index.php). The online hourly data is also freely available through FMI's open data system. The history data can also be downloaded by anyone without restrictions from HSY's open data system or from the Finnish air quality portal. The air quality information is also sent to the EU register by FMI. (Niemi 25 April 2016.)

The air quality in the Helsinki metropolitan area is mainly deteriorated by exhaust gases and street dust from traffic, residential wood burning in fire places as well as long-range transported pollutants from other countries of Europe. HSY's official website shows general information on air quality and air protection as well as the online concentrations on an hourly basis. (Niemi 25 April 2016.)

HSY also measures basic meteorological parameters with one-minute resolution (e.g. temperature, wind speed and direction, atmospheric pressure, relative humidity, radiation and rainfall) at a few measurement stations. However, the quality control of meteorological measurements is not very high since they are mainly used to support general interpretation of air quality monitoring results. The high quality meteorological observations are provided by the FMI in the Helsinki metropolitan area in the whole country. (Niemi 25 April 2016.)

Pegasor

The data originates from measuring devices, like sensors that measure air particle concentration (nanoparticles) and are able to produce a long time series of the classifying data. The data is collected at a fixed point. Once installed, the monitoring station produces data around the clock, typically every five minutes, but can do even once every second. (Saukko 2 May 2016.)

Vaisala

Vaisala doesn't specialize in air quality and it produces weather measuring devices and sells them to the Finnish Meteorological Institute (FMI). It doesn't have its own data source. (Laakso & al. 18 May 2016.)

Tampere University of Technology (TUT)

The data of the Tampere University of Technology (TUT) come from measurement campaigns conducted for vehicle and energy production emissions. TUT is currently collecting its data by using long-time series of nanoparticle data from HSY's monitoring station located on the Mäkelänkatu street in Helsinki. In addition, they have a mobile laboratory which can be used to study air quality in different environments. (Rönkkö 7 June 2016.)

5.2.4 The Contents of the Data

The Finnish Meteorological Institute (FMI)

There are two types of model data within the Cityzer system. First, some model data are inputted into the Cityzer system. These data are not included in the Cityzer output data and are generally invisible to the Cityzer clients. Second, the Cityzer system includes models that generate model data. These data are the principal outputs of the Cityzer system. The input model data include predicted weather conditions for the next 24 h from the HIRLAM numerical weather prediction (NWP) model, regional air quality prediction from the SILAM model and predicted precipitation type. (Nousiainen 20 April 2016.)

They contain the observed precipitation as a function of location, measured every 5 minutes. The radars measure the power of the reflected microwave radiation transmitted by the radar, as a function of time and location, which is then converted into the rainfall

intensity and further into the rainfall accumulation. In addition, most of the radars in the FMI radar network measure also polarization of the reflected microwave radiation. This provides additional information about the precipitation and allows identification of radar echoes that are not due to rainfall but, for example, of non-meteorological origin such as birds. The polarimetric quantities measured also help in identifying the precipitation type, either snow, rain or hail. The lightning measurements are obtained by the Nordic lightning information system (NORDLIS), in part operated by the FMI. This network records individual lightning strikes, including their location and time, using triangulation of the low-frequency radio waves emitted by the lightning. The FMI has augmented the national network with high-frequency sensors to infer additional information about the lightning activity. (Nousiainen 20 April 2016.)

The predicted weather conditions include the meteorological data, such as temperature, pressure, relative humidity and wind, close the surface as a function of time and place for the next 24 hours. These data are obtained from the HIRLAM NWP model run at the FMI, and used by the FMI-ENFUSER model to generate the Cityzer air quality forecasts. Likewise, the regional air-quality predictions produced by the SILAM model are used as input for the FMI-ENFUSER model and its air quality forecast for the Cityzer domain. The format of the HIRLAM NWP data, SILAM data and ENFUSER now- and forecasts is in the netCDF format. The format of the weather radar data is HDF5 and of the lightning data is UALF. (Nousiainen 20 April 2016.)

Metadata is data about the data. Much of the metadata is generated along with the data it describes, by the same processes that generate the data described, but some is generated through other mechanisms. For example, the Cityzer system will generate metadata to describe the system performance. In addition, the Cityzer system performs certain diagnostics on the data within the system and may add diagnostic metadata on it. The generation, structure and content of the metadata are still being planned and cannot be commented on in more detail for now. (Nousiainen 20 April 2016.)

Helsinki Region Environmental Services Authority (HSY)

Data is stored in HSY's central server and most of it is also, distributed through to the Finnish air quality portal maintained by to the Finnish Meteorological Institute, and it transfers it as open data. Sometimes it changes the resolution from an hour-resolution to a 15-minute resolution. All the Finnish provinces send their air quality data to FMI through the air quality portal, and FMI publishes the results online as a map in the Finnish air quality portal “Ilmanlaatu.fi” website. The typical format of data is .csv. (Niemi 25 April 2016.)

Pegasor

The weather data is limited and contains information about temperature and humidity. The particle concentration signal is the primary signal, which is accompanied by metadata. The metadata is received from the sensors and has data about, possible error notifications, quality assurance data, with which the service can confirm that the data is reliable, and the sensor is working properly. The data from the device complies with the Modbus protocol and after being stored in the cloud database service, the format can be changed to xml, .csv, or .txt format. (Saukko 2 May 2016.)

Tampere University of Technology (TUT)

The data of TUT consists of three elements. One of them is detailed information on the amount and characteristics of particle emissions of individual engines, vehicles and power plants. Particles are most component of air pollution, and in urban areas those particles are released into the air due to traffic and in some countries also by energy production. “Detailed” refers to the fact that TUT has data on particle size distribution, particle number emissions, particle volatility and electric charge carried by particles. The second is a time series of particle concentrations and size distributions as well as some other things in the atmosphere. The university has data not only from Helsinki, but also from Tampere. a partner university in Beijing has a collaborator, which can provide some data for further use. The third is information on the function of new measurement instruments. The format of the raw data depends on the instrument used in the measurement, e.g. .txt and .dat are typical formats. (Rönkkö 7 June 2016.)

Below is a summary of the format of the Cityzer data.

Table 3. Data Format of the Cityzer data

Partner	Data / Process	Data Format
FMI	HIRLAM NWP data SILAM data ENFUSER now- and forecasts	netCDF
	Weather radar data	HDF5
	Lightning data	UALF
HSY	Air quality portal	.csv
Pegasor	Modbus protocol	xml, .csv or .txt
TUT	Instruments	.txt and .dat

5.2.5 Data Storage

The Finnish Meteorological Institute (FMI)

The weather observation data maintained by FMI are stored in databases, in backed up hard drives. The data includes the measurements at ground stations, radio sounding and measurements of solar radiation. From the measurements, climatic parameters are also calculated and stored in a separate climate database. (Nousiainen 20 April 2016.)

The model data storage practices produced by FMI vary. For example, data of regional weather forecasts, produced by the weather-predicting HIRLAM model, are stored in the tape archive of European Center for Medium-Range Weather Forecasts (ECMWF). On the other hand, the data of the HARMONIE model, used for modelling smaller-scale weather phenomena, are stored in the archives at the FMI, in a backed up hard drive. The measurement data of the weather radar network of FMI is stored centrally in backed up hard drives. (Nousiainen 20 April 2016.)

In addition, the Finnish Meteorological Institute is involved in many satellite and space research projects. Regarding them the data storage policies vary. For example, the data from instruments sent to different planets are stored centrally by the international PDF (Planetary Data System) organization. The FMI Arctic Center in Sodankylä, in turn, receives and stores data from many satellite instruments that measure the Earth's atmosphere. There is a huge amount of other data with no centralized documentation, in some cases, there is no documentation. The data may be completely under individual research-

er's responsibility, for example backed up in an external hard drive. (Nousiainen 20 April 2016.)

Helsinki Region Environmental Services Authority (HSY)

Data is stored in HSY's central server and most of it is also distributed to the Finnish air quality portal maintained by the Finnish Meteorological Institute, and they transfer it as open data. All the Finnish provinces send their air quality data to FMI, and FMI publishes the results online as a map in the Finnish air quality portal "Ilmanlaatu.fi" website. (Niemi 25 April 2016.)

Pegasor

The data which is produced is transmitted to the cloud services. The data can be fetched straight from the verifiable cloud service application for further analysis. This is what the company offers other companies. In a nutshell the core business of Pegasor is the production of high quality sensors and provide data transfer and storage in the cloud, from where the data can be used for further analysis or data fusion with other sources. (Saukko 2 May 2016.)

Tampere University of Technology (TUT)

The data is only in TUT's servers and it is not public. (Rönkkö 7 June 2016.)

5.2.6 Data Classification

The Finnish Meteorological Institute (FMI)

The data can be classified into measurement and modelling data. The output model data include precipitation nowcasts (0-3 h), lightning nowcasts (0-3 h) and air quality forecasts (0-24 h). The precipitation nowcast is computed with the RAVAKE nowcasting model. The model analyses the development of observed precipitation from the latest three radar composites (e.g., 0, 5 and 10 minutes ago) and extrapolates it 0-3 h into the future. The nowcast thus contains data on modeled precipitation in the Cityzer domain as a function of time and place. (Nousiainen 20 April 2016.)

The lightning nowcast is a new, as of yet non-existing product to be developed and implemented within the Cityzer project. It will be generated analogously to the precipitation

nowcast. Recent data on lightning strikes will be converted into lightning density maps (number of lightning strikes per area) within the last 5, 10 and 15 minutes, from which the lightning density can be extrapolated into the future. (Nousiainen 20 April 2016.)

The air quality forecast is computed with the FMI-ENFUSER model that combines statistical air quality modelling (Land Use Regression), dispersion modelling techniques (Gaussian plume) and information fusion algorithms. Based on the air quality observations (HSY network), meteorological data (HIRLAM NWP model), estimated precipitation (RAVAKE model within the Cityzer system), and data on the aerosol background, including long-range transport (SILAM), high-resolution air-quality forecast is computed for the Cityzer domain, covering always the next 24 hours. The forecast includes data on PM2.5, PM10, O_3 and NO_2 as a function of time and location at the surface within the domain. The impact of precipitation on air quality will be accounted for probabilistically, allowing for probabilistic air quality predictions (e.g., probability that NO_2 concentration exceeds certain threshold). (Nousiainen 20 April 2016.)

Helsinki Region Environmental Services Authority (HSY)

Concentration values of pollutants are usually converted to Finnish air quality index values when online results are presented to public. The index classifies pollutant concentrations into five air quality classes varying from good to very poor. The Finnish air quality index is based on hourly mean concentration values. (Niemi 25 April 2016.)

The air quality parameters are mostly measured with continuous online methods. However, some parameters are also measured from samples in laboratory analyses. For instance, heavy metals and benzo(a)pyrene are measured from PM10 samples with 24 h time resolution. Volatile organic components (VOC) are measured from passive samplers with 2-week time resolution. (Niemi 25 April 2016.)

Pegasor

The primary data stream is the active surface area of the particles and measuring how much particulate matter there is in the air. The data can be classified e.g. by the metadata, consisting of the time, location and temperature and humidity information. The other classifications are the prevailing weather conditions and wind direction with which we can get information the emission sources around the measuring point. The information provides better classification to get extended information from the surroundings and e.g. the sources around the measurement site. (Saukko 2 May 2016.)

Vaisala

The data can be classified as open or closed data or free or data with a cost. (Laakso & al. 18 May 2016.)

Tampere University of Technology (TUT)

The air quality data of cities can be classified according to the parameters measured or by the environments where the measurement has been made. Vehicle / engine emission data can be classified also by parameters measured by technology used in test vehicles / engine. Power plant emission data are not very large but they can be classified based on the type of power production. (Rönkkö 7 June 2016.)

5.3 Visualization of Results

Below is a table summarizing the results: the data source, contents, storage and classification of each company and organization.

Table 4. A summary of the Cityzer project (Roles and Data Source)

Partner	Role	Data Source	Data Contents	Data Storage	Data Classification
FMI	Data Provider, System Orchestrator, main founder and owner of the Cityzer project	In situ, nowcast, weather forecast models (FMI-ENFUSER, SILAM, RAVAKE), weather radar measurements, lightning measurements and air quality measurements	observational data, model data, meteorological data (temperature, pressure, relative humidity and wind), input data, model-generated data, metadata	databases (backed up hard drives), climatic database (climatic parameters), regional weather forecast data stored in tape archive of ECMWF, small-scale weather phenomena data stored in FMI's archives, satellite data from instruments stored in PDF organization.	precipitation nowcast (0-3 h), lightning nowcast (0-3 h) and air quality forecast (0-24 h), derministic nowcast (precipitation in 3h time), probabilistic nowcast (precipitation > 10 mm/h with 50% probability), amount and type of precipitation
HSY	Data Provider, Data Consumer, "testbed" (= tests the sensor network provided by Pegasor and air quality fusion model of FMI, air quality monitoring	11 monitoring stations in traffic sites, small house areas, harbors as well as urban and rural background sites, Online monitors measure gaseous (NO_2, NO, O_3, SO_2) and particulate pollutants, which record data in a one-minute resolution measured air quality	data regarding street dust and small particles from exhaust gases (NO_2, SO_2, black carbon), particulate mass as m^3 of air and exhaust and road dust emissions	HSY's central server ➔ air quality portal ➔ sent to FMI as open data	Finnish air quality index classification of pollutant concentrations into five air quality classes (good – very poor), based on hourly mean concentration values, online methods, laboratory samples (heavy metals and benzo pyrene - PM10 samples with 24 h time resolution, (VOC) – passive samplers with a 2-week time resolution.
Pegasor	Data Provider, provides air quality information on particles and their behavior to others companies / orgnizations	sensors → air particle concentration (nanoparticles)	air quality emission sources (exhaust fumes from traffic and street dust), information about temperature and humidity, metadata (quality assurance data)	data stored in cloud services ➔ used for further analysis	amount of particulate matter in the air, metadata (time, location, temperature and humidity), weather conditions, wind direction, data around the clock, every 5 minutes
Vaisala	Device Provider, produces devices and sells them to other companies (FMI)	no data sources, produce weather measuring devices and them to FMI			open or closed data, free or data with a cost

TUT	Data Provider, supports the research on air quality sensor networks (experiments, traffic emission modelling, conducting, calibration and comparison studies for particle sensors, workshops)	measurement campaigns (vehicle and energy production emissions), nanoparticle data collected from HSY's monitoring station, mobile laboratory	detailed information on the amount and characteristics of particle emissions (individual engines, vehicles and power plants), a time series of particle concentrations and size distributions, information on the function of measurement instruments	TUT's servers, data is not public	parameters measured or by the environments where the measurement has been made, parameters measured by technologies used in test vehicles / engine, power production (power plant emission data)

Below is a diagram representing the roles of each partner in a reference architecture model.

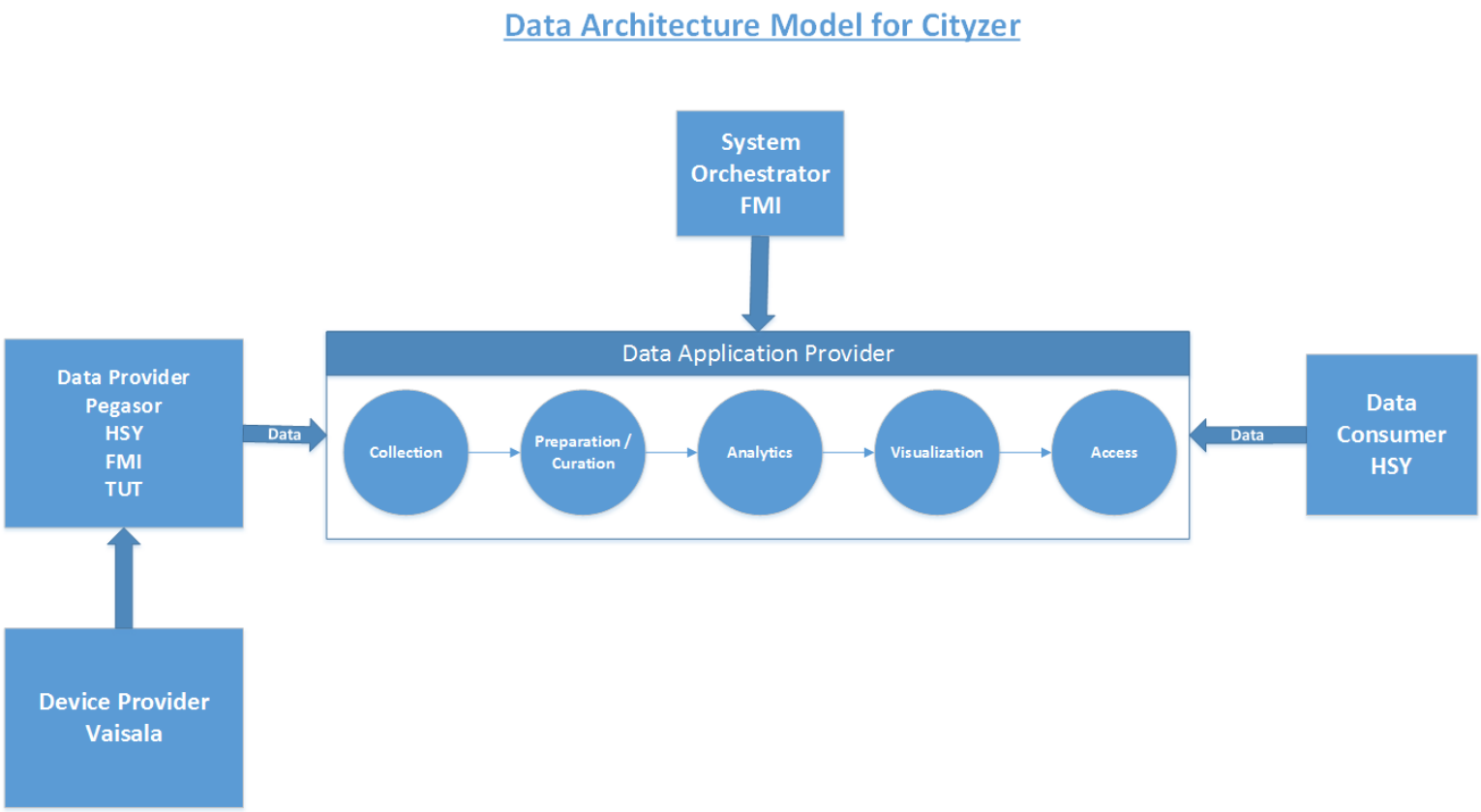

Figure 6. Data Architecture Model for Cityzer.

6 Conclusion

From the results, it can be concluded that the role of most of the partners of the Cityzer projects are Data Providers, in other words, they provide data to each other and their customers, so that they can benefit from it. Some of the partners use the services of the other and provide support for collecting the data. They collect the air quality and/ or weather data from measuring devices (e.g. measuring stations and Vaisala's devices) and either publish them online or store them in cloud services or in a database or servers. Some of the partners conduct the research and try to cover new ways of measuring weather and air quality. The data can be classified by time resolution or weather parameters. Most of the data is private, except for HSY's and FMI's are public, but the data of HSY can only be obtained as open data from FMI's air quality portal. The format of the data is mostly in .csv or .txt, except for the data of FMI collected from the forecast models, which are in different formats.

The Cityzer data flow diagram (appendix 1) depicts the whole data flow process of the Cityzer project. From the diagram, it can be concluded that the whole project works as a chain and the partners are "connected" to each other, as they co-operate with each other and are in some ways dependent on each other. For example, FMI is the project owner and makes sure the chain is running smoothly. HSY collects data using its own devices and later on, send the data to FMI through the portal, to publish as online data. HSY also tests Pegasor's sensors, which collect air quality data, specifically nanoparticle data. TUT, on the other hand, collects air quality data by using HSY's monitoring station. FMI collects weather data using Vaisala's own weather devices. They all need to work with each other, in order for the project to succeed. It is unclear whether the roles of the Cityzer partners will change in the future, as FMI seems to have a strong role as the System Orchestrator and HSY as one of the main Data Providers. It is possible that these two organizations can also have the second role as the Data Consumers, as some of the other partners (e.g. TUT) uses HSY's monitoring stations to collect data and HSY could use TUT's data to its benefit. Vaisala expressed their wish to become the System Orchestrator in the future, as they wanted to see the big picture in the whole "chain". That way they could create new services and bring value to the "chain". Pegasor's and TUT's role might remain the same.

Out of all the data providers, HSY, FMI and TUT are in a position to manage, analyze and share the data further to other possible service providers like mobile application providers. These providers could be Emtele and Sasken. These companies are also Cityzer partners and specialize in mobile products and remote intelligent cabinet for sensor networks.

These data companies can send the weather and air quality data to Emtele, Sasken and other companies, which can create Weather Forecast API to use in the existing or newly-launched weather applications for consumers, and also inform about air pollution. Another option is cloud storage, which Pegasor specializes in.

Interviewing was chosen as the method of research and it proved to be the best method for collecting information, interviews are a far more personal form of research than questionnaires. One benefit is that the interviewer has the opportunity to probe or ask follow up questions. Interviews are generally easier for the respondent, as they can freely express themselves; their thoughts and opinions. This proved to be the case also in this thesis, as all the companies and organizations were keen on finding out the data sources of the other organizations, and the only way to get reliable information was by interviewing all of the partners. They were able to provide some background information on the topics and even ask questions regarding the objectives of the thesis from the interviewer.

The duration of the whole research process was four months, from February to June 2016. It was mainly conducted for the personal use of the partners to use for the Cityzer project, however, it was a good learning experience as well. This project taught how to search for information using different sources (e.g. online material from the Internet, literature review), conducting interviews and learning about the different interview methods and analyzing results and visualizing them into a table and diagram. It also gave the opportunity to become involved in a major national project and become acquainted with the different organizations and companies that were involved in this Cityzer project. It also taught a lot about data collection and data analysis and gave an insight on the air quality and weather situation in Helsinki. That is the reason why this topic was chosen, as it provided an opportunity to learn about something new, as weather and air quality was not such a familiar topic.

The most challenging aspect of the thesis was to, first understand the objectives of the Cityzer project and figure out who the thesis can relate and benefit the project and later schedule interviews, as most of the partners had conflicting schedules.

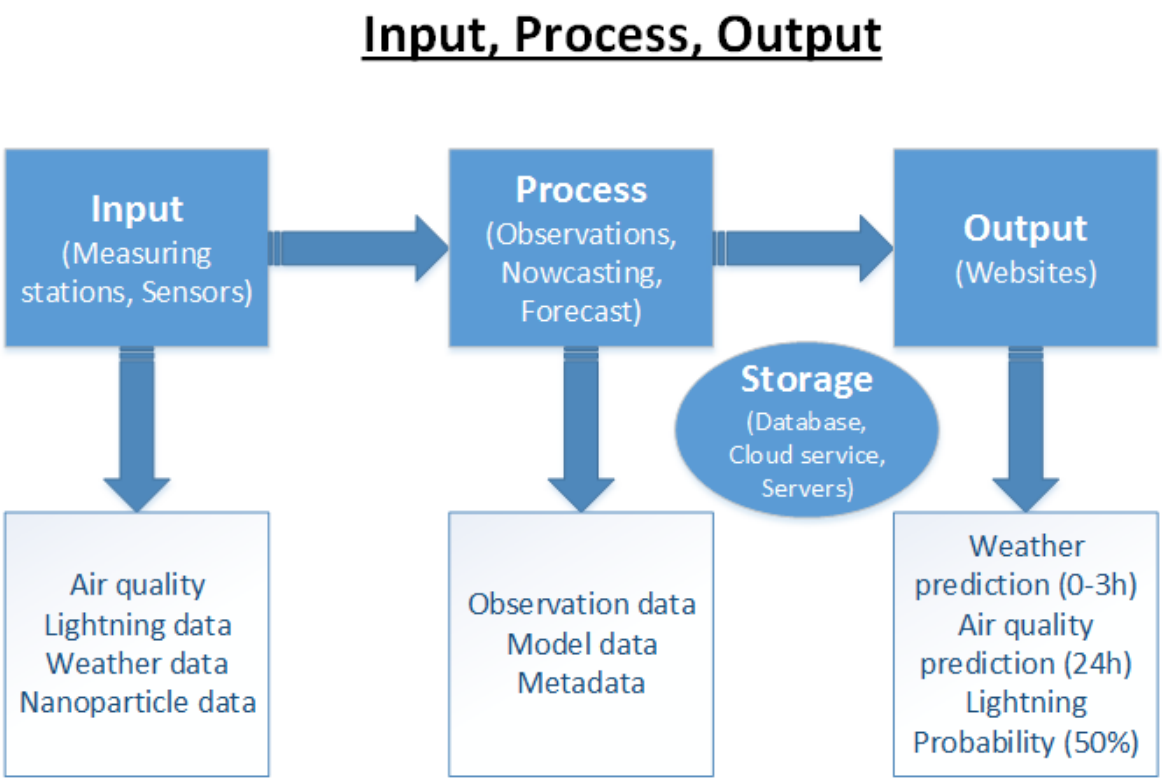

Figure 7. The input data, process and output data of Cityzer

The diagram above picturizes the data flow and explains what happens to the data in the whole process. The input data (air quality, weather data and nanoparticle data) are measured by measuring stations, which measure the particles in the air (fine and thoracic particle mas PM10 and PM2.5 and particles such as nitrogen oxides NOx and sulphur oxide SO_2 and sensors. The data are transmitted through an air quality portal and published as online results in websites (FMI.fi, HSY.fi and Ilmanlaatu.fi).

7 Utilization of the results

This thesis project concentrated on gathering information on the weather and air quality data of the partners of the Cityzer project. The goal was to figure out what the data included and where it originated from. This provided the companies with valuable information, which they could later for the project. The thesis was an analysis of the data in the Cityzer network: mapping the data source, data format and metadata, including a framework to categorize this data for further use. (Cityzer official project plan.)

The companies of the CITYZER consortium will use the results of the project in growing their business in Finland and internationally. It is expected that companies can renew their business by expanding to new sectors or market environments. For export markets, total solutions are targeted, that can be exported as turnkey deliveries after adaptation to market specific conditions. Also for individual products and services, i.e., parts of the system, growing business opportunities arise as well. To facilitate successful business creation, knowhow on market structures and business models suitable for the most important market areas (China, India, Brazil) will be elaborated. (Cityzer official project plan.)

The thesis will in the future be used by HAAGA-HELIA, Tampere University of Technology and the leading partners The Finnish Meteorological Institute, Helsinki Region Environmental Services Authority, Vaisala, Pegasor, Emtele, Sasken, INNO-W and CLIC Innovation to examine all the data sources of the Helsinki region and use them Cityzer project for concepting new types of services. In the future, they will use the data to determine if it is free or can be sold to other parties and, thus, create business models.

8 Summary

The topic of this thesis is “Weather and air quality data of Helsinki” and the main objective was researching, analyzing and classifying the contents and of the weather and air quality data for the Cityzer project. The final objective was to map and understand the data and the business ecosystem around it, and then classify the data and paint a picture of the whole ecosystem around the data. The aim was to work with the weather companies and partners, such as Vaisala, Pegasor, The Finnish Meteorological Institute, Helsinki Region Environmental Services Authority and Tampere University of Technology, and analyse how they collect data and where they are stored. This thesis was an analysis of the data source, form and origin. This theory is divided into five parts: “Air Quality and Weather”, “Helsinki’s situation”, “Big Data”, “The Internet of Things” and “Data, Information, Knowledge and Value”.

It looks deeper into the weather and air quality situation in the Finnish capital, Helsinki, and studied the factors that affect weather and air quality, how weather and air quality can be measured and the current situation and problems of the air quality in Helsinki. It also gives a brief insight on Big Data and its architecture and also the Internet of Things. The main research question is “What kind of data is available for analysing weather and air quality in the Helsinki region?” and sub questions backing up the main question are the data provided by the Cityzer project partners, the weather and air quality data: the origin of the data, their contents and storage place and how they were classified. The different partners and companies were interviewed to understand their role in the project to collect answers to the question mentioned above.

From the interviews, it was concluded that most of the partners provided the data to other partners and third parties, such as FMI, HSY, TUT and Pegaosr. Vaisala manufactured devices and sold them to the other organizations, so that they could collect the data via the devices. FMI is also the founder and system orchestrator of the Cityzer project. Most of the air quality data were collected by monitoring stations and sensors, and the biggest concern was pollutants. The data contains detailed information on air quality emission sources and particle emissions and also metadata. The data are stored in hard drives, cloud services or in servers. In HSY’s case, it sends the data to FMI through an air quality portal and FMI, in turn, publishes them online. The data are classified by different models, nowcasting, Finnish air quality index or weather parameters. There were other partners of the project that we not interviewed, such as INNO-W, Emtele, Sasken and CLIC Innova-

tion, as they didn’t deal with the data at all but were involved with the business side of the project.

The mission of the Cityzer project is to develop new digital services and products to support decision making processes related to weather and air quality in cities. The long-term goal of the project is to issue early warnings and forecasts, which would prevent weather-related accidents, lessen human distress and costs from weather-related damage and poor air quality, and generally improve the resilience and safety of the society. The data provided by project services could be used to guide rescue services and plan preventive actions and design secure solutions in city planning. This thesis will provide the partners with useful information on their data, which they can demonstrate to outside parties. The CityzerDEMO is a pilot demonstrating the Cityzer ecosystem in the Helsinki Metropolitan area. CityzerDEMO is currently being planned and implemented, and is expected to be opened for public in late 2017 or early 2018.

Bibliography

BC Air Quality. Factors Affecting Air Quality. Air Quality. Air Quality 101. What is Air Quality? URL: http://www.bcairquality.ca/101/air-quality-factors.html. Last accessed: 12.03.2016.

BC Air Quality. How We Measure Air Quality. Air Quality. Air Quality 101. URL: http://www.bcairquality.ca/101/measure-air-quality.html. Last accessed: 12.03.2016.

Bhuiyan, F. Vaisala's weather instruments. Photograph

Robert, L. 2010. What is Information? EcoLabs. Image online. URL: https://ecolabsblog.wordpress.com/2010/11/29/data-information-knowledge-and-wisdom. Last accessed: 06.04.2016.

City of Helsinki. 2015. Factors affecting air quality. Housing and Environment. Environmental protection and oversight. Air Quality and Noise. Factors affecting air quality. URL: http://www.hel.fi/www/Helsinki/en/housing/environmental/air-noise/factors-air/. Last accessed: 19.03.2016.

Climate and Weather. 2014. Climate & Weather. Characteristics of world weather and climate. URL: http://www.climateandweather.net/world-weather/factors-that-influence-weather.html. Last accessed: 12.03.2016.

Dunbar, I. 2013. The Infogeenering Model. Ian Dunbar's Weblog. Image online. URL: https://ecolabsblog.com/2010/11/29/data-information-knowledge-and-wisdom/. Last accessed: 18.06.2016.

Finnish Meteorological Institute. About Us. URL: http://en.ilmatieteenlaitos.fi/about-us. Last accessed: 11.06.2016.

Green, C. 2014. Big data and Mapping – a potent combination. Information Age. IT Management. Strategy and Innovation. 28 July. URL: http://www.information-age.com/it-management/strategy-and-innovation/123458292/big-data-and-mapping-potent-combination. Last accessed: 25.02.2016.

HAAGA-HELIA. 2013. Degree regulations. URL: http://www.haaga-helia.fi/en/students guide/welcome-haaga-helia/degree-regulations. Last accessed: 06.4.2016.

HSY. About HSY. URL: https://www.hsy.fi/en/abouthsy/Pages/default.aspx. Last accessed: 11.06.2016.

HSY. 2015. Air pollutants and emission sources. Image online. URL: https://www.hsy.fi/en/residents/theairyoubreathe/information-on-breathing-air/Pages/default.aspx. Last accessed: 18.06.2016.

HSY. 2015. Information on breathing air. The air you breathe. URL: https://www.hsy.fi/en/residents/theairyoubreathe/information-on-breathing-air/Pages/default.aspx. Last accessed: 24.05.2016.

HSY. Measured components in 2016. Image online. HSY image collection.

HSY. 2015. Monitoring stations in Helsinki. Image online. URL: https://www.hsy.fi/en/residents/theairyoubreathe/monitoring-stations-helsinki-metropolitan-area/Pages/default.aspx. Last accessed: 18.06.2016.

Infogeneering. 2016. The Differences Between Data, Information and Knowledge. URL: http://www.infogineering.net/data-information-knowledge.htm. Last accessed: 06.04.2016.

IoT Agenda. TechTarget. 2014. Internet of Things (IoT). Definition. URL: http://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT. Last accessed: 06.04.2016.

Kemppainen, S., Kuha, J., Arvola, L & Karjalainen, J. 2015. Automated Water Quality and Weather Monitoring in Lakes Jyväsjärvi, Konnevesi and Vanajavesi. MMEA / Deliverable D.WP1.V.DATA.10 Jyväskylä 2015. CLEEN Cluster for Energy and Environment. Paijanne. URL: http://www.paijanne.org/media/MMEA_deliverable_JYU_D_WP1_V_DATA_10.pdf. Last accessed: 20.03.2016.

Laakso, M. 18 May 2016. Product Line Manager. Vaisala. Interview. Vantaa.

Mahendra, A. 2016. How Internet of Things will help in weather forecasting. IoT Worm. February 6. URL: http://iotworm.com/internet-of-things-technology-weather-forecasting. Last accessed: 06.04.2016.

Mahendra, A. Remote station illustration. IoT Worm. Image online. URL: http://iotworm.com/internet-of-things-technology-weather-forecasting/. Last accessed: 18.06.2016.

Morgan, J. 2014. A Simple Explanation of 'The Internet of Things'. Forbes. Leadership. May 13. URL: http://www.forbes.com/sites/jacobmorgan/2014/05/13/simple-explanation-internet-things-that-anyone-can-understand/#13d0f8876828. Last accessed: 06.04.2016.

Niemi, J. 25 April 2016. Air pollution expert. Helsinki. Helsinki Region Environmental Services Authority. Interview. Helsinki.

Nousiainen, T. 20 April 2016. Senior researcher. Finnish Meteorological Institute. Interview. Helsinki.

Ojanperä, O. 18 May 2016. Product Manager. Vaisala. Interview. Vantaa.

OTEC. Data, Information, Knowledge, and Wisdom. URL: http://otec.uoregon.edu/data-wisdom.htm. Last accessed: 06.04.2016.

Pegasor. Company. URL: http://pegasor.fi/en/company/. Last accessed: 11.06.2016.

Pritzker, P. & May, W. 2015a. NIST Big Data Interoperability, Framework: Volume 1, Definitions. Final version 1. pp. 4-12. NIST Big Data Public Working Group. Definitions and Taxonomies Subgroup. URL: http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf. Last accessed: 25.02.2016.

Pritzker, P. & May, W. 2015b. NIST Big Data Interoperability, Framework: Volume 2, Big Data Taxonomies. Final version 1. pp. 11-24. NIST Big Data Public Working Group. Definitions and Taxonomies Subgroup. URL: http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-2.pdf. Last accessed: 14.03.2016.

Prtizker, P. & May, W. 2015b. NIST Big Data Reference Architecture. NIST Big Data Interoperability, Framework: Volume 2, Big Data Taxonomies. Image online. URL: http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-2.pdf. Last accessed: 18.06.2016.

Roponen, J. 2016. Analytics Architect. Big Data and the Internet of Things, Business opportunities, technologies and examples. IBM. Lecture slides. Big Data course. Spring 2016. February 17. Helsinki.

Rönkkö, T. 7 June 2016. Research Manager. Tampere University of Technology. Email. Tampere.

SAS Institue Inc. Big Data Analytics. What it is and why it matters. Insights. Analytics. URL: http://www.sas.com/en_us/insights/analytics/big-data-analytics.html. Last accessed: 06.03.2016.

Saukko, E. 2 May 2016. Senior R&D Engineer. Pegasor. Interview. Tampere.

Ståhle, L. 18 May 2016. Senior Systems Architect. Vaisala. Interview. Vantaa.

Tampere University of Technology. About TUT. URL: http://www.tut.fi/en/about-tut/. Last accessed: 11.06.2016.

Vaisala. 2014. Vaisala Automatic Weather Station AWS310 / An Innovative Weather Solution for All Needs. URL: http://www.vaisala.com/Vaisala%20Documents/WEA-MET-AWS310-Brochure-210x280-B211290EN-C-LOW-v2.pdf. Last accessed: 24.05.2016.

Walliman, N. 2011. Research Methods: the basics. Routledge. New York.

Wikipedia. Vaisala. URL: https://en.wikipedia.org/wiki/Vaisala. Last accessed: 11.06.2016.

Weather for Schools. Collecting Weather Data. Data Collection. Collecting the Data. URL: http://www.weatherforschools.me.uk/html/collectingdata.html. Last accessed: 12.03.2016.

Weather for schools. ICT and the Weather. Data Collection. Using ICT. URL: http://www.weatherforschools.me.uk/html/collectingdata.html. Last accessed: 12.03.2016.

Appendices

Appendix 1. Cityzer Data Flow Diagram

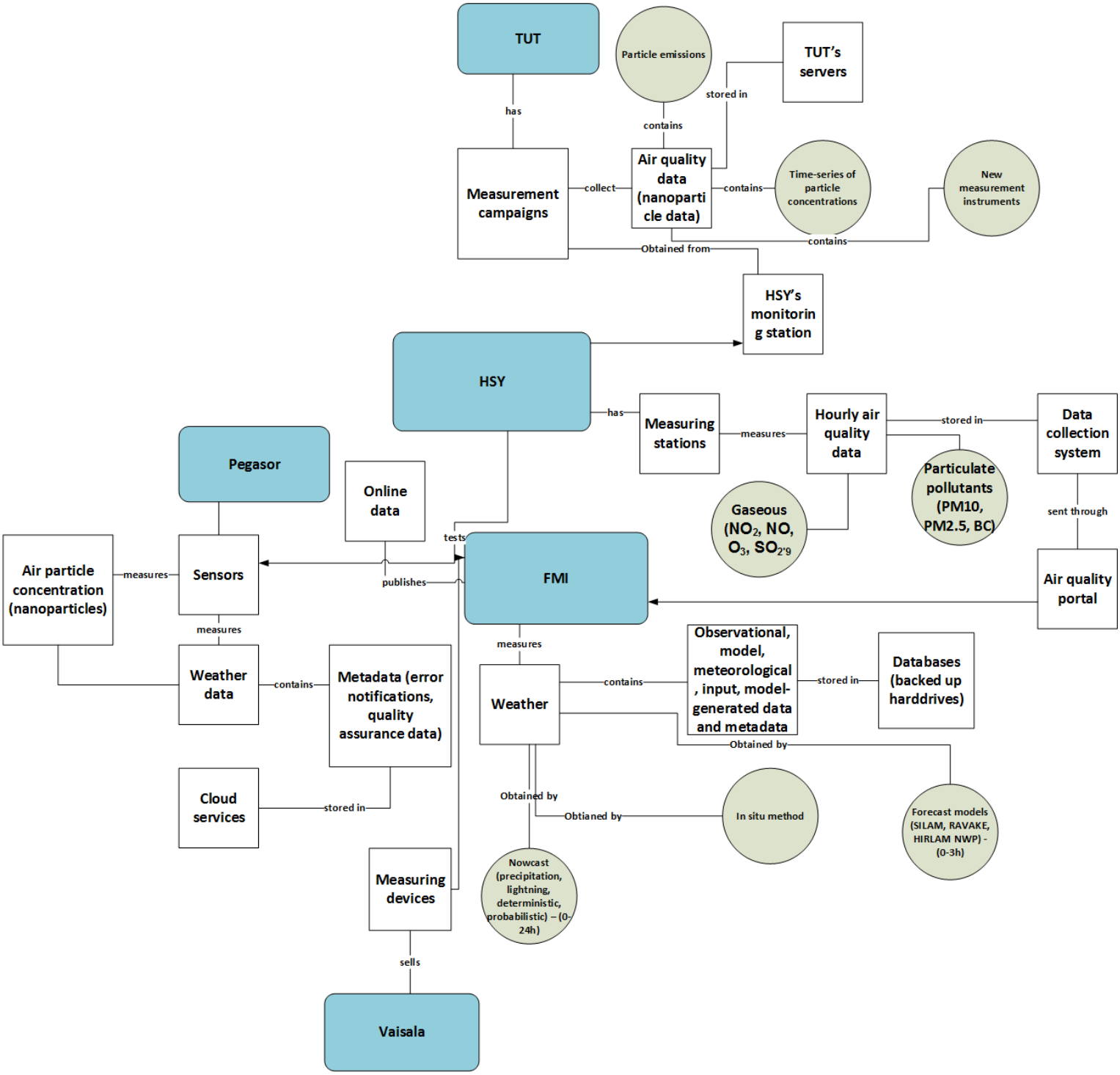

Figure 8. Cityzer Data Flow Diagram