EPS - PROJECT

TITLE: DIGITAL TWINS II

STUDENTS:

Wout Desart

Nastia Cherednichenko

Pedro Boueri Palmeira

Alexandru-Daniel Nicolae

Daniel Sánchez Molina

SUPERVISORS:

Félix Ruiz Gorrindo

Nora Martínez Antúnez

DATE: 07.06.2021

Second semester 2020-2021

FAMILY NAME: Desart	**FIRST NAME: Wout**
HOME UNIVERSITY: Artesis Plantijn Hogeschool Antwerpen	
SPECIALITY: Automation (PLC programming)	

FAMILY NAME: Cherednichenko	**FIRST NAME: Anastasia**
HOME UNIVERSITY: Technical University of Koshice	
SPECIALITY: Business Informatics	

FAMILY NAME: Palmeira	**FIRST NAME: Pedro**
HOME UNIVERSITY: Universidade Federal de Itajubá (UNIFEI)	
SPECIALITY: Industrial Engineering	

FAMILY NAME: Nicolae	**FIRST NAME: Alexandru-Daniel**
HOME UNIVERSITY: Politehnica University of Bucharest	
SPECIALITY: Mechanical Engineering	

FAMILY NAME: Sánchez Molina	**FIRST NAME: Daniel**
HOME UNIVERSITY: Universitat Politècnica de Catalunya	
SPECIALITY: Mechanical Engineering	

Abstract

Digital Twins have been around since the early 2000s, but it has only been until now that they started to be affordable thanks to the Internet of Things. In the realm of smart cities, a Digital Twin is a virtual model of a city, a replica of the physical world, which are rapidly becoming indispensable tools to visualize the pulse of the city in real time with layered data sources of buildings, urban infrastructure, utilities, businesses, movement of people and vehicles. The advantages of implementing this concept is that it significantly increases the city's stability. Testing in a virtual model helps prevent emergencies, properly allocate resources that reduces costs and the chances of failure in the real world.

This project is a continuation of the last year's theoretical study Digital Twins I and its aim is to continue the research about Digital City Twins and explore the Big Data from the city sensors of Vilanova i la Geltrú. A group of five international students, led by the company Neapolis, are working on transforming the city into a smart one within the summer semester of the academic year 2020-2021. In the process, we studied scientific articles, consulted with university professors from different countries (Spain, Belgium, Brazil), contacted IT and Data Security companies to obtain the necessary information. The report provides a study of practical examples using Digital Twins around the world, their impact on the city improvement, comparison of different platforms and software for developing Digital Twins and the reasoned choice of the best option for use in the next part of the project. Furthermore, it describes Information Infrastructure of Digital Cities, Big Data Management, Data Security and the implementation of Digital Twins in Vilanova i la Geltrú. The Big Data received from the city authorities was read and analyzed in the data part with necessary conclusions. This project made a great contribution to the further development of the Digital Twins for Vilanova i la Geltrú and will simplify the practical implementation for our followers of the next EPS project.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

1. Introduction

1.1. Project Description

In order to carry out the goals under the EPS (European project semester) project, we have been grouped into a team of five international students. The members of the team for the Digital Twins II project are:

Name	School	Country	Field of studies
Alexandru-Daniel Nicolae	Universitatea Politecnica din Bucuresti	Romania	Mechanical Engineering
Wout Desart	Artesis Plantijn Hogeschool Antwerpen	Belgium	Automation
Anastasia Cherednichenko	Technical University of Koshice	Ukraine	Business Informatics
Pedro Palmeira	Universidade Federal de Itajubá	Brasil	Industrial Engineering
Daniel Sánchez Molina	Universitat Politècnica de Catalunya	Spain	Mechanical Engineering

Table 1 Group Introduction

Nowadays, new technologies are part of our daily lives making life simpler and easier. This is the reason why Neapolis makes a study about the Digital Twins concept. They want to create a new model of lifestyle in Vilanova i la Geltrú by making it a smart city, connecting the real world with a virtual model of the city. The development of this concept is through a program based on Internet of things (IoT) powered by sensors which are capable of detecting problems, preventing problems and “costs, trail and shape” predicted outcomes in 3D representations of the city.

Continuing with the project carried out by the European Project Semester (EPS) students, from last semester, DIGITAL TWINS I, where they conducted the in-depth research on the concept of Smart Cities and how to implement it in Vilanova i la Geltrú. The objective of our project is to continue with Their research and theoretically implement the digital twins concept in the coastal city, by proposing and applying a model to manage Big Data.

Neàpolis is a place that connects business, technology, coworking and design. It is a center of innovation, entrepreneurship and creativity which belongs to the Vilanova i la Geltrú City Council. The company has already participated in various previous projects with the Escola Politècnica Superior d'Enginyeria de Vilanova i la Geltrú (EPSEVG), with the aim of helping the city of Vilanova i la Geltrú and Neàpolis to grow multiple aspects.

DIGITAL TWINS II is the continuation of the last project carried out by Neàpolis and EPSVG in the European Project Semester, in which the implementation of a Smart City is sought through the Digital Twins concept. The international team is made up of engineers covering different fields such as manufacturing, mechanical, electromechanical automation and business informatics. It also includes a supervisor from the university, Nora Martínez, who is helping in the project management aspect and a supervisor from Neàpolis, Félix Ruiz, to help with more specific points about the project. The project will be carried out through group and autonomous work, with weekly meetings with students and others with supervisors. The project follows a carefully planned path, obtained through a gantt chart. The different goals and tasks are tracked and managed in the Trello application.

DIGITAL TWINS II continues with the same final objective as the previous project, which is to convert Vilanova y la Geltrú into a Smart City based on the concept of digital twins, making the life of the citizens of the city much easier and more comfortable.

1.2. About the company "Neàpolis"

Neàpolis is a public agency that promotes innovation in technology, designs, universities, coworking and entrepreneurship. **Neàpolis** offers training, premises and support to technological companies and business start-ups. They also provide more learning opportunities through a community which aspires to entrepreneurial excellence. **Neàpolis** belongs to the City Council of Vilanova i la Geltrú. At the same time technological, creative and business projects are being developed in just one place. Neàpolis also leads courses for companies and professionals.

Figure 1 Neàpolis

The company Neàpolis is in the southeast of Vilanova i la Geltrú. The company building is located in the industrial area. Furthermore, Neàpolis is based in the immediate vicinity of the UPC University site. This short distance makes it easy to establish connections and interactions between the academic and business worlds. In addition, it should be added that the location of Neàpolis is close to the railway station and has good road connections. For this reason, the building of Neàpolis is very easy to reach.

The current district can be defined as an industrial area. The district provides two large parking lots, located in the south and in the middle of the district. It has six bus stops and a train station, which is located in the southwestern part of the district. Furthermore, there are supermarket chains in the south of the district. Besides this, the entire neighborhood is characterized by a high degree of mobility. The district contains streets which are open to traffic. Therefore, the district is easily accessible.

Neàpolis consist in:

- A park of Technological and Business Innovation.
- A meeting point for companies, professionals and entrepreneurs in the ICT and Media sectors.
- An ecosystem where environments of innovation and research, training and technological dissemination and business activity coexist.

Figure 2 Neàpolis main center

Neàpolis is also the main center of audiovisual production in the south of Barcelona. It hosts a film and TV school that offers degrees in film, performing arts, digital cinematography and transmedia content. It hosts leading laboratories and research teams from the Universitat Politècnica de Catalunya. It acts as a node of local telecommunications and is the venue for major activities and events related to the fields of innovation and technology from the Global Game Jam to scientific congresses, conferences and seminars.

Mission:

- To promote economic activity and the co-creation of innovative solutions with a social impact in the sectors of technology , creativity and communication , in order to turn Vilanova i la Geltrú into an innovative intermediate city of reference.

Vision:

- Neàpolis wants to become a reference for society in the development of innovative solutions that:
    - They cover the social needs of the citizens and the city .
    - Promote social technology , creativity and business drive .
    - They are generated from the collaboration between the key agents of the territory.

Values:

- **Innovation:** Always ready to find creative solutions that respond to the needs of the citizens of Vilanova i la Geltrú.
- **Creativity:** Creativity is a key element in making innovation possible. Promote the hybridization of different professional fields, to share seemingly opposite perspectives and to have, whenever possible, creative agents (designers, artists…).
- **Social technology:** The technology put at the service of society, provides great value as it improves the efficiency of management and provides new solutions. The social impact of technology brings benefits to the majority and reduces the unwanted consequences.
- **Collaboration:** Neàpolis is a collaborator that encourages the meeting and interaction between the key agents of the territory so that they find shared solutions.
- **Uncertainty Management:** Neàpolis encourages the deployment of innovative solutions that adapt to this uncertainty or take into account the complexity inherent in each social need.

1.3. Project goals and objectives

Since this project is the continuation of the previous EPS project - Digital Twins Ⅰ, it has related goals and one important purpose - to help turn Vilanova i la Geltrú into a Smart City based on the concept of digital twins making the life of the citizens of the city much easier and more comfortable.

The outcome of this project is to delve into the study of the Digital Twins, manage and read the Big Data from the city and to make it useful to the Neàpolis and the City Council. In other words, everything to get closer to developing the practical application of the Digital Twins of Vilanova i la Geltrú.

The main objectives to carry out the project are:

- Study the characteristics of the previous EPS project about Digital Twins.
- Study existing models to manage Big Data (including algorithms, 2D representations, decision making, etc.).
- Propose a model to manage the Big Data for Vilanova i la Geltrú.
- Propose an example of Big Data of Vilanova i la Geltrú, including different parameters (traffic, humidity, etc.). Some of the data can be simulated, if it is not possible to get real data.
- Apply the proposed model (data management, machine learning startup) in point 3 to the proposed Big Data of Vilanova i la Geltrú.
- Get 2D representations of the proposed Big Data of Vilanova i la Geltrú.
- Decision-making based on the application and costs of the proposed methodology to Big Data.
- Pros & Cons analysis of the methodology used and of the obtained results.
- Propose the model to apply in other countries, study the benefits and difficulties and propose further steps for implementation.

These objectives will vary as the project develops due to the limited knowledge we have on the subject due to our background.

2. About the previous project Digital Twins Ⅰ

The previous project Digital Twins Ⅰ was carried out by five international students and was the first step to create a digital version for Vilanova i la Geltrú. Their final work describes the main concepts associated with Digital Twins, history and current situation; advantages/disadvantages of implementing Digital Twins to smart cities; general algorithms and used sensors, different processes after data collection and some other things. In addition, they mentioned planification: how the next company can take steps towards a smart city concept and project cost.

1. Hence, within the framework of the Digital Twins Ⅰ project, the following goals were achieved:
2. Study the characteristics of Neàpolis and the city of Vilanova i la Geltrú
3. Study the current situation of the Digital Twins concept around the world
4. Describe all the advantages and propose the implementation of the Digital Twins concept in Vilanova i la Geltrú
5. Propose a plan of phases to implement the concept
6. Calculate the cost of these phases
7. Propose an occupation plan, in order the work be made by people and companies of the territory
8. Introduce the concept hybridization, in the sense that different professions collaborate generating new transverse professions
9. Propose a methodology to package the concept of implementation of Digital Twins in Vilanova i la Geltrú, in order it can be exported to other places/countries
10. Investigate the future evolution of the Digital Twins concept around the world

Based on this more theoretical previous part, we continued our investigation and delved deeper into the details of creating Digital Twins and turning ordinary cities into smart ones.

3. Digital Twins cities in the world

3.1. Reasons why cities need Digital Twins

To understand the impact of a smart city we must first understand the process behind cites, a city is not an automated system, it is a very complex process that can be compared to the system of a living organism. All these systems will vary from day to day, which makes it extremely difficult to make predictions and correct decisions. These factors will be enormously dependent on underlying factors and will therefore fluctuate. As the current problem, the world is under the massive attack of the covid 19 pandemic, some domains such as education, transportation and entertainment will be attacked heavily. By using digital 3d representations, also known as digital twins, these factors can be tested and influenced in order to make the right decisions. Not only this, smart cities and digital twins will positively influence the quality of life of their inhabitants. A smart city will therefore boost the well-being of its residents on key final points. The main focus will be on the sustainability of the city, in terms of economy and ecology, and on improving the delivery of services to its inhabitants. The implementation of smart cities or digital twins will therefore benefit almost every area. Below, one can find the main advantages of implementing the Digital twin cocept, which will also apply to the city of Vilanova. The biggest keypoints in our project will be traffic and ecology.

Governance	City
1. Sharing trusted data across city departments and the entire design and build ecosystem to support infrastructure development from planning to operations 2. Combining disparate data sources 3. Systematic tracking, control and visualization of changes 4. Providing analytics to key decision makers 5. Involvement of all stakeholders and the public 6. Operational cost savings	1. Energy efficiencies 2. Improved sustainability 3. Positive impact on economic growth 4. Safer communities 5. Safety is improved 6. New economic development opportunities 7. Increased digital justice 8. Reduced ecological footprint 9. Improved infrastructure

Table 2 Advantages SmartCity Governance/City

3.2. Different Smart Cities

3.2.1. Comparison between Smart cities

3.2.1.1. Singapore

Virtual Singapore, created by the National Research Foundation (NRF), is a government department that offers 3D semantic modelling, in which the meaning of data can be related to the real world, displaying land attributes or the characteristics of different forms of transport, or the components of buildings and infrastructures.

"Dassault Systèmes worked with the National Foundation of Singapore to create a complete virtual twin representation of the city using the 3DEXPERIENCity® solution. By combining geometric, topological and environmental data with information on everything from climate to traffic patterns, Singapore can run simulations and virtual tests to understand and develop solutions to urban planning challenges.

Creating these virtual models allows researchers and city planners to apply different data sets to scenarios to understand their potential effect. Something that is particularly relevant during times of global disruption. For example, by including data in a city's virtual twin about the distribution of hospital beds and protective personal equipment (PPE), the virtual twin could have guided emergency responders to transport patients to hospitals with available beds, while helping city officials direct excess supplies to where they were needed most."

Uses of virtual Singapore:

In the gardens of the bay

- Cloud Forest replicates the weather of cool and tropical mountains.
- Flower dome replicates dry mediterreane climate with 7 differents gardens inside there.

Light and temperature

- The virtual singapore can calculate percentage of lights and shadow per building based on the position of the sun.

Inside the virtual singapore there is the Bishon District which is an accurate visual representation of the physical district from 2D geospatial layers to 3D city with semantic information, it can provide information about address, level of the apartment, post code, time exposed to the sunlight per day and also the material of the building. Information such as Area and volume of the building is available and Geo-localized information connected to the geographic information system, providing information about parking lot and trees.

Virtual Singapore provides sustainable transportation by autonomous vehicles using 3D simulation models and polygons as well as the street furniture for driverless navigation on the roads. Residents can also check the availability of AV's in the area

Smart walking

- Cietzen use virtual Singapore for smart walking to their destination from MRT stations through overhead bridges to void-decks and amenities in their neighbourhood solar panel production

- Virtual Singapore is able to check in real time the amount of electricity generated (kwh) by solar Panel voltaic. With the features previously presented the city of singapore is very close to its objectives which are:
- Use AI to address major challenges that affect society and industry AI can be used to increase traffic throughput during peak hour, for example, or to address healthcare challenges that are to come with an ageing population. Healthcare is currently both a knowledge and human-touch-intensive industry. Coupled with the progress in the digitization of Singapore's healthcare over the years, AI could be significantly applicable for safeguarding the health of Singaporeans. AI could play a big role in supporting prevention, diagnosis, treatment plans, medication management, precision medicine and drug creation. Healthcare manpower, augmented with AI tools, could better address increased healthcare demands in the future.

- Invest in deep capabilities to catch the next wave of scientific innovation These may include next-generation "explainable" AI systems exhibiting more humanlike learning abilities, as well as adjacent technologies such as computing architectures (integrating software, firmware and hardware) and cognitive science. The NRF Fellowship and Investigatorship schemes will be deployed to support such scientific activities. Local talent in AI will also be trained through their involvement in the development of these deep capabilities in AI.

- Broaden adoption and use of AI and machine learning within industry AI Singapore will work with companies to use AI to raise productivity, create new products, and translate and commercialize solutions from labs to the market. NRF aims to deliver 100 meaningful AI projects and proofs-of-concept to solve real-world problems quickly for end-users. There is particular potential in the sectors of finance, healthcare and city management solutions, which the program will start with.

3.2.1.2. Amaravati

Amaravati is a small village located in the guntur district, the region is located between the two largest regions of Andhra Pradesh; guntur and Vijayawada. Amaravati is located over a piece of land of 200 square kilometers consisting of many small villages, barren landscapes and the river Krishna in the north.

Figure 3 Amaravati Structural plan

The city gets rated by the Indian Green Building Council (IGBC) Green cities Rating, therefore they got a rating "Platinum", which is the highest possible rating a city can obtain. The city has a smart integrated infrastructure plan with the follow varied features:

<u>Traffic and Transportation:</u>

Public transportation is made so that everyone has access to the necessary means of transportation, eliminating the need to use motor vehicles. This will remove most of the risk factors associated with transport in India and will improve air quality and overall pollution.

"The effective policies and proposals to be implemented to achieve the set goal are to promote shared public transport by the year 2050 keeping the targets at 70% across the city and within the CBD its 80%. At the initial period BRT is planned which eventually will be developed to MRT. A high speed rail corridor to mark the pace of growth of the city is proposed. The city's transport is designed on smart lines in such a way that it takes five minutes to reach essentials, ten minutes to any commodities fifteen minutes for any public transport." (Ghadei, 2017)

Wastewater Management:

The waste water management system will be constructed in a very robust and controlled manner. Bad systems can lead to huge health problems. This is why we are going for a robust system, with separate pipes and the necessary smart implementations to perform the necessary measurements, leaks and maintenance.

StormWater Management:

The city is located on very flat land next to the river Krishna. From research and the necessary smart data obtained from this river, one can predict what and when the river will overflow its banks. With the necessary adjustments, one obtains a very low flood profile. By using the excess water present, they can start implementing the following applications.

“They are- Rainwater harvesting through which 40% runoff volume is taken care of and helps in percolation.

- Grass swale is a technique used to decrease stormwater runoff by attenuating or treating it.
- Green roof is another practise that reduces runoff volume by 45-55 %. It also reduces the urban heat, creates recreation with aesthetic beauty.
- Porous pavement method is adopted which allows storm water to drain through them and is absorbed by the native soil beneath or is detained temporarily. It is estimated to control runoff by 45 %.
- Detention ponds at planned places to decrease the peak time rise and are expected to reduce runoff by 50%.
- Rain gardens to infiltrate the rain water or flood water temporarily.

  In this way loss due to property destruction and health costs to be spent on treating water borne diseases can be avoided.” (Ghadei, 2017)

Power:

The city will use a highly advanced electricity grid, to ensure continuous supply of power. Smart grid, smart metering, underground cabling, LED lighting and many other applications will form the basis of this infrastructure. All this data will in turn be collected in SCADA called supervisory control and data acquisition center. ICT will play a very big role in this environment. Continuous monitoring and flexibility will be present. ”It would then serve as a platform of growth for tackling several smart applications. Andhra Pradesh optical fibre grid is worth mentioning here which is implemented with the vision of establishing a high scalable network providing broadband connectivity of up to 20 mbps to households and 10 mbps to enterprises by 2018. The ICT plan will realise the dream of digital AP with the combined efforts of GoI and the private sector.” (Ghadei, 2017)

Safety and Security:

Security will be guaranteed by continuous monitoring of the streets by cctv cameras. This will not only eliminate the basic dangers but also try to predict and prevent man-made disasters.

Conclusion:

With all these smart implementations, Amaravati constitutes India's first green city. It will give a progressive future to thousands of households and be an inspiration to India and the rest of the world.

3.2.1.3. Sant Cugat

Sant Cugat del Vallès is a municipality in Catalonia, located in the region of Vallès Occidental, extending to the north-western slopes of the Collserola mountain range, until the beginning of the Prelitoral depression.

It has grown a lot in recent years, due to its high birth rate, the highest in Spain in 2004 and the arrival of newcomers, this is the reason why their average age is 36 years. The city council is very involved in turning the small city into a smart and green city, for this they have the objective of promoting economic, social and environmental sustainability for people, entities and companies in order to improve the quality of life. The actions that the city has carried out are the following:

Outdoor Lighting Plan:

Remote control of the light schedule and intensity in the 20,000 light points through 210 control panels and 130 light regulators.

The entire commercial axis, provided with 340 light points, has motion sensors, so that in the absence of pedestrians the intensity of the light is 20 percent until it increases to 100% in the presence of pedestrians.

Many of the incidents can be resolved from the control center or, where appropriate, are referred to technicians, equipped with advanced technological systems, who come to the site in the shortest possible time.

With this lighting plan, a 30% reduction in the energy consumed and a 76% reduction in light pollution is achieved, which entails a reduction in emissions of 1,049 tons of CO_2 / year.

Traffic:

The center and commercial area of Sant Cugat is pedestrianized. In terms of traffic, the inverted priority is established (limited access for loading and unloading at specific times exclusively for previously identified and authorized vehicles). The 25 access points, previously equipped with bollards, are now controlled with cameras and radars.

In addition, there are 33 regulated crossings with traffic lights throughout the city. Any incident is detected and corrected from the control center.

Noise:

Through the Smart Lighting Noise system, developed with Mediaurban, it helps to detect and, where appropriate, reduce the acoustic level in crowded areas at night.

It consists of intelligent sensors incorporated into some streetlights that when they detect that the decibel level allowed by the municipal ordinance has been exceeded for more than a minute, they emit flashes. In this way, customers are alerted to lower their voice. The establishment is also warned.

The table below shows the indicators that are regulated in the different cities. These factors are determined and coordinated by the big data obtained from the city. The indicators of Vilanova i la geltru are theoretically determined and would be needed in the city. [1]

	INDICATORS										
CITIES	Impact of weather conditions on city functioning	Very high impact of weather conditions on city functioning	High population density	High pollution	Accessibility problems	Noise problems	Criminality	Garbage problems	Parking problems	Class extremes	Diverse languages, races, cultures, religions
Singapore	☒	☒	☒	☒	☒	☒	☒	☐	☒	☒	☒
Amavarati (India)	☒	☒	☒	☒	☐	☒	☒	☒	☐	☐	☐
Barcelona	☒	☐	☒	☒	☒	☒	☒	☒	☒	☒	☒
Sant Cugat	☒	☐	☐	☐	☐	☒	☐	☒	☐	☐	☐
Vilanova i la Geltrú	☒	☐	☐	☐	☒	☒	☐	☒	☐	☐	☐

Table 3 Comparison SmartCities

The following table shows the differences between Smart Cities such as Sant Cugat, Amaravati and Singapore. This will be one of the first comparisons and a guideline for further research.

	SINGAPORE	AMARAVATI (India)	SANT CUGAT (Spain)
Roadmap	Offer not only maps and land data, but also other real-time dynamics. Increase traffic throughput during peak hour	Remove most of the risk factors associated with transport in India	Guided vehicle parking
City-planning	Predict how the opening of a new plant will affect the environment or how the construction of a stadium will change the situation on the roads	+	-
Demographics	+	Predict population growth, which should reach 11 million by 2025	-

Health care	By including data about the distribution of hospital beds and protective personal equipment, guide emergency responders to transport patients to hospitals with available beds, while helping city officials direct excess supplies to where they were needed most	-	-
Using street light sensors	+	-	Street lightings sensitive to presence; pedestrian crossing equipped with lighting sensors
Using air quality sensors	+	+	-
Using sunlight sensors	Predict amount of sunlight	Distributed solar power generation	-
Using moisture sensors	+	+	Irrigation system dependent of humidity levels
Using container sensors	-	Detailed waste management	Intelligent garbage refuse collection point
Using water quality sensors	-	Sewage treatment	-

Conclusion	The largest project for the development of an information 3D model worth $ 73 million (2014 - 2020). The 3D map will include not only buildings and roads, but even doors, windows, park benches and lampposts	Amaravati is supposed to have blue-green spaces on more than 30 percent of the city's area. Renewable power and sustainability are the biggest priorities. Various measures are being taken for green cover protection and water conservation	Smart street idea was applied to the whole city. It allowed to reduce environmental and economic bill by 30% after only one year

Table 4 Comparison SmartCities 2

3.3. Results and examples of real cases

Stockholm (Sweden) uses its OpenCities Planner for streaming information models of territory, containing large amounts of data overlay for quick and easy information exchange. The virtual twin allows solving strategic and local tasks.

result

It is possible to predict and visualize various scenarios for the behavior of the public at large city events, thanks to which the best way to organize the entrances and exits in the areas of the greatest concentration of people, thereby increasing safety.

The digital 3D model of **Antwerp (Belgium)**, launched in 2018, combines data on the city's noise pollution with information on air quality and traffic conditions. The virtual twin allows accessing up-to-date data from various city services in real time, makes accurate forecasts for the future and makes optimal decisions in the shortest possible time.

result

At the touch of a button, it is possible to predict various scenarios for the development of urban systems and analyze their impact on each other.

The digital twin of the city of **Rennes (France)** is based on the same technology as the twin of Singapore - 3DExperienceCity by Dassault Systèmes. Virtual Rennes aims at

result

If the city is considering developing real estate in a specific area, citizens could potentially see the project on a public 3D map. In addition, the model makes it possible to predict the level

effective interaction between the municipality and residents.		of noise or air pollution that will result from construction. Thus, residents will be able to form their opinion about the project and communicate it to the city.
The digital twin of **Rotterdam (Netherlands)** is designed to solve urgent problems for the city, such as water transport management, waste disposal, and fire-fighting operations.	result	It is possible to track the intensity of traffic on roads and waterways, optimize the opening/closing of bridges and redirect traffic flows in real time.
The virtual model of **Helsinki (Finland)** is a digitized version of a massive 1915 physical model. It was created on the basis of more than 42 thousand images taken from an altitude of 1.2 km. The city's digital twin is being developed using the CityGML standard in the new seaside district of Kalasatama. The virtual model is open to any user, which allows the city to introduce and improve new technology.	result	For example, it is possible to generate an atlas for calculating and visualizing the solar energy potential of a city or to perform a digital analysis of the effect of wind on high-rise buildings.
The digital concept of **Newcastle (UK)** is the first city model designed to protect the city's infrastructure and residents from natural disasters. Authorities decided to develop a virtual model of the city after heavy rain caused £ 8 million in damage in 2012.	result	Technology enables cities around the world to respond quickly and efficiently to natural disasters - rising sea levels, weather anomalies, drought and energy shortages
The authorities of the city of **Takamatsu (Japan)** have implemented 2 digital twins: a digital twin for monitoring and preventing emergencies (in particular, floods; based on the collection and analysis of data from water level sensors) and a digital twin for the city's tourist attraction (monitoring of the movement of rented bicycles was introduced).	result	The digital twin of emergencies provides real-time monitoring of the flood risk of each of the city's districts, also monitoring the status of shelters for city residents by collecting information from humidity sensors and electricity consumption in each of the shelters (mobile application available to every resident). Comparison of the geo-track of the movement of rented bicycles with data about the tourist (age, gender, citizenship, etc.) allows creating a map of the most important tourist locations in the city, planning and evaluating the effectiveness of

marketing campaigns to promote tourist attractions.

4. Algorithms

The creation of an algorithm capable of processing all the data and choosing the useful information generated by the sensors is one of the most important and difficult tasks of this project.

To start with the creation of the algorithm we have to understand its definition very well, which consists of a set of defined and unambiguous, ordered and finite instructions or rules that allows, typically, to solve a problem, perform a computation, process data and carry out other tasks or activities, or otherwise, given an initial state and an input, by following the successive steps, a final state is reached and a solution is obtained.

A digital twin for a smart city is a set of fictitious actions, exactly similar to reality, which are generated from the data provided by a large number of sensors based on real-time data of different characteristics. Creating the digital twin of a smart city requires a good understanding of the processes. Currently, the team does not have too much exact information about the sensors that will be used in the city, which is why we are going to use real (old)data for simulation of the algorithm and in the future we will exchange them for real data.

First, a process map needs to be created with all the data input and operations. The core module of the DT is a simulation that is modelled based on the process map. The process map provides the requirements for the data points needed for the model. The key requirements for the implementation are (1) definition of the action scenario to be carried out, (2) operational data capture and analysis towards the identification of key parameters (3) creation of the digital twin with the integration of key motion parameters and operational constraints, modelling the behaviour of the physical assets, and (4) simulation of the smart city process and its optimization according to a set of optimization constraints.

A simulation algorithm has different stages. First, initial data needs to be provided and the time period needs to be defined.The simulation runs and stops after each time period and waits for “new data input”. Then, after each simulation run, the simulation stops and asks two questions: 1) is there new data input? 2) does user want to review the last performed data? From our answer, the simulation algorithm can branch out in different ways:

1. "NO" new data input, and the user does "NOT" want to review the last performed data: the simulation moves to the next time period using the previous data input. The data review will not be executed either.
2. "YES" new data input, and the user does "NOT" want to review the data : the simulation moves to the next time period using the new data input. The data review will not be executed.
3. "NO" new data input, and user "WANT" to review the data: the simulation moves to the next time period using the previous data input. In the future, depending on the data of each sensor, the simulation algorithm will be able to optimize and improve the results in the case that the data reviewed is not positive.
4. New data input "YES" and user "WANTS" to review the data: the simulation moves to the next time period using the new data input. The data review segment will be executed as per aforementioned (3.) segment.

Once we have the basis of how our algorithm will be, one of the things to highlight to do in the future is how the algorithm will be able to simulate various actions from real life in the case that the data it got was not positive, and choose the best option to improve the city. This step will be simulated in the future when the project obtains more information about their tasks.

A problem that arises in the creation of our simulation algorithm of digital twins for the city of Vilanova i la Geltrú, is whether we must create an algorithm for each type of sensor which is capable of optimizing the results only in a specific area, or the creation of a great algorithm that is capable of obtaining and analyzing all kinds of data and finding solutions for all areas.

On the one hand, the option of a simulation algorithm for each type of data, gives us the facility to correct the errors that we have made in the creation of the algorithm and the ability to manipulate it. On the other hand, a general simulation algorithm means that we only have to work on one, which makes it a more complicated and extensive task since the probability of failure is higher. Our first idea is to create a simulation algorithm at the beginning for the different types of big data provided by the city council and in the future to group them all in a single general simulation algorithm. This process may vary at any time in the future because the project does not yet know what type of data it is going to face exactly.

5. Existing platforms/software for developing Digital Twins

5.1. SmartWorldPro

Amaravati has had software developed in-house. The software currently in use is SmartWorldPro by Cityzenith. The city invested no less than 6.5 billion dollars in the development of their new smart city.

The software is based on a single concept. Unifying all data generated in the city. It uses a crystal-clear 3D visualisation system that allows certain data to be instantly viewed and compared. Based on this, the software is made to easily design and handle one or more buildings. The software is almost limitless in size, of both users and buildings. In addition to this immense size, SamrtWorldPro is equipped with an SDK (software development kit) that allows users to integrate 3rd party applications very easily. The software is available to all, and customised budgets make it possible for both large and small companies to step into this "revolution".

Hybrid cloud/on-premise approach, offers the users both the advantages of cloud storage and local storage of project data. Because the software is created with the idea of being as open-source as possible, it creates a unique outcome. There are possibilities to redirect the software APIs and workflow engines to other cloud-based programs. This makes it possible to link the software with Microsoft Azure, among other things. Currently, SmartWorldPro uses AWS for its flexibility and is secured by its robust disaster recovery system.

In contrast to the open source approach, the SmartWorldPro is very cybersecure. The data is stored at all times in S3 encrypted rest banks or within an encrypted-at-rest database. In addition, the rights to this data will be limited to certain users within a certain project and the data will be securely transferred only when absolutely necessary.

Furthermore, both this software and other SmartCity programmes focus on a very important growth in efficiency. As mentioned earlier, by 2050 we will have to accommodate an immense increase in population, compared to 10,000 new cities. To make this growth possible and as efficient as possible, we need to make use of Smart Cities. This software also aims at an efficiency increase of almost 25%.

The current software is working on projects all over the world. They are located in Chicago, Oxford,Riyadh, San Francisco, Aberdeen, Kuwait, Rhode Island, Orlando and Amaravati.

5.2. 3DEXPERIENCity®

3DEXPERIENCity was the software used to do the 3D simulation of singapore. The software is powered by Dassault systemes who is very renowned for CAD softwares as Solidworks.

The Dassault Systemes’ 3DEXPERIENCE platform connects an entire ecosystem on a single cloud-based platform, from which you can leverage the 3DEXPERIENCE virtual twin to bring together the real and virtual worlds. This enables your business to not only better respond to disruptions as they arise, but also collaborate and innovate more effectively with stakeholders to rethink and reshape your business so it can thrive in the future. [3]

Dassault Systèmes Brands develop applications that form the building blocks of the **3D**EXPERIENCE® platform. [2]

- Social and Collaborative (ENOVIA, CENTRICPLM, 3DEXCITE)
- 3D Modeling (CATIA, BIOVIA, GEOVIA, SOLIDWORKS)
- Simulation (3D VIA, SIMULIA, DELMIA)
- Information Intelligence (EXALEAD, NETVIBES)
- The applications are used to create Industry Solution Experiences (ISE)

The software 3DEXPERIENCity can have a wide range of applications but the most important for this projects are:

- Visual experimentation
- Test-bedding
- Decision making
- Research and development

One very interesting way to transform 2D maps into 3D interactive city is shown in the following link and can be very usefull for the next EPS group. (https://www.youtube.com/watch?v=mhqh7u3zHpY)

5.3. Microsoft & Bentley Systems alliance

Bentley Systems is a leader in engineering software for professionals to design, build, operate and maintain critical infrastructure such as road and rail networks, and public works and utilities, and will collaborate to develop new smart city solutions.

Along with Microsoft, Bentley Systems will explore opportunities for digital twins in urban planning and citizen engagement for cities around the world. This includes city planning projects in Dublin and the renovation of FC Barcelona’s Camp Nou stadium.

Bentley’s expertise is in aggregation, alignment, and 3D visualization of complex engineering data, as well as the tracking of change. The front-end of the digital twin runs in a standard web browser and relies on Bentley’s iTwin Viewer for 3D visualization. The back-end cloud services run in Azure and data from multiple data sources. The back-end ingests and normalizes data from 2D computer aided design (CAD) files, 3D building information models (BIM), and geographical information systems (GIS). Projects typically start with the model of their physical environment and then federate data from IoT devices to the model. Azure Digital Twins abstracts the IoT layer and provides APIs, which make it easy for developers to integrate live data.

A significant benefit of the iTwin-Azure Digital Twin integration is the simplification and time saved by developers by automatically generating from the iTwin the digital twin definition language (DTDL) that is required to create the spatial intelligence graph in Azure Digital Twins.

The figure below describes this configuration schematically and gives an overview of the hand-in-hand cooperation of the two programs.

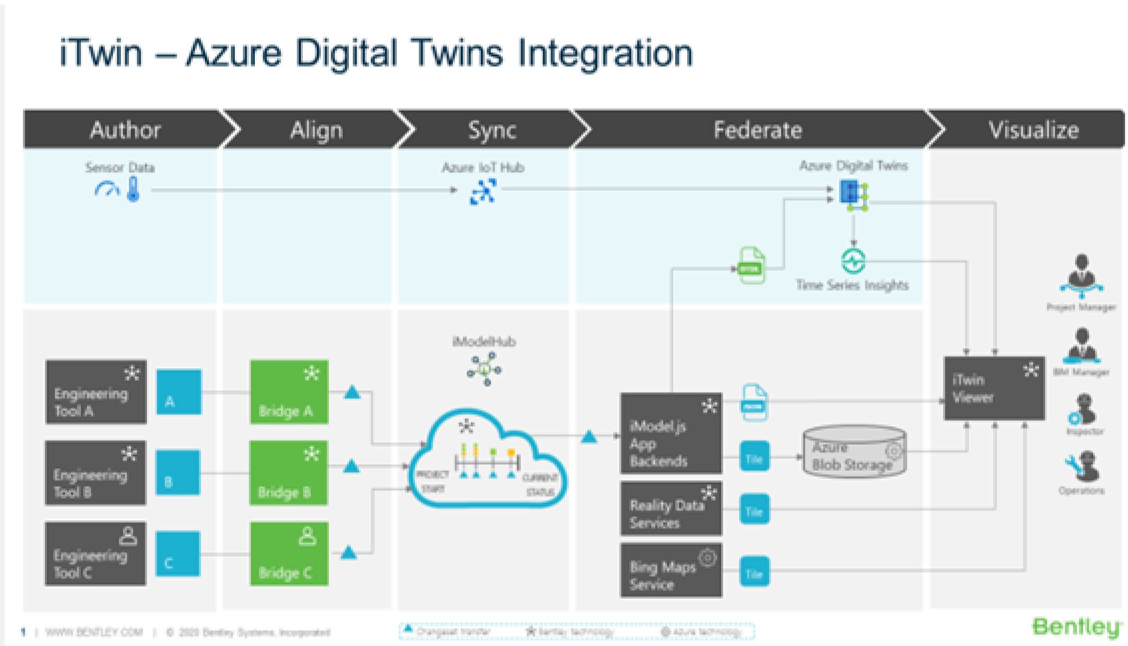


Figure 4 Azure Digital Twin Integration + Bentley systems

The explanation of the different processes of the picture will be developed in section 13.

5.4. Microsoft Azure Digital Twin

Azure Digital Twins is an Internet of Things (IoT) platform that enables you to create a digital representation of real-world things, places, business processes, and people. Gain insights that help drive better products, optimize operations and costs, and create breakthrough customer experience

The software was fully developed in December 2020 and is a hub for different types of data. It offers the advantage of ready to use building blocks and is a important tool that converts all type of data collected from sensors and other sources of data for DTDL (Digital Twin Digital Language).[4]

The uses of Azure Digital Twin are many, some of them are tailored for the project goals:

- Iot solution: more connected to other devices
- Spaces and facilities: one example is smart building
- Products: Manufacturing goods
- Process: Can be applied for supply chain processes in order to have a real time view about how the process is going.

As a SaaS (Software as Service) the Azure Digital Twin brings to the developers some facilities that makes the process of building a digital twin less complex. The developers do not need anymore to develop the devices, they can use existing building blocks and use the time to describe the places, relationships, improve security and project a scalable project.

The software is tailor fit for digital twin and internet of things, using digital twin digital language (DTDL) and pre ontologies being fed by partners. All this with the aim to spend less time “gluing” all these pieces and spend more time layering these developers' value on 3D and 4D visualizations, developing artificial intelligence to the project differenciationing their models.

One very powerful tool of azure digital twin is that the developer do not have to write the full code, it can be taken from CRM data and from azure Iot hub (where the Iot data are stored), the developer just have to write a code that will put all this information together,saving time for activities that add value to the project. That’s the reason why this software was selected by the group to develop this project.

5.4.1. How to get access to software

The UPC is a member of the Microsoft Azure Dev Tools for Teaching initiative (formerly Microsoft Imagine). The benefits of the program do not change, it allows teachers and students with current enrollment, both in the classroom and at home, the full range of development and design tools, server products, operating systems, applications and Microsoft information libraries, so they can use them for research and teaching. In our case we used one of the newest softwares from Microsoft Azure which is Microsoft Azure Digital Twins.

To access the software you must follow the steps indicated on the university website, https://serveistic.upc.edu/ca/distsoft/faq/microsoft-imagine/acord-upc-microsoft-imagine , in our case we had a problem during the access. Being a new software, the servers had not yet been updated where they indicated that the members of the upc could use microsoft azure digital twins, finally after contacting the microsoft azure technical service they solved the problem.

5.4.2. Price of the software

The product's website contains a section where you can check the price of the different operations that appear on the Digital Twin. The price depends on your territorial area among other factors, in our case Vilanova i la Geltrú is located in Western Europe, the price is as follows:

Explore pricing options

Apply filters to customize pricing options to your needs.

Region: West Europe

Currency: Euro (€)

	Price
Message	**€1.097** per million messages
Operation	**€2.741** per million operations
Query unit	**€0.549** per million query units

Figure 5 Price of Azure Digital Twin in West Europe

5.5. Software comparison table

SmartWorldPro VS 3DEXPERIENCity®

SmartWorldPro	3DEXPERIENCity
Cloud storage and local storage	Cloud base platform
Crystal-clear 3D visualization system	Visual experimentaion
Easily design and handle one or more buildings	Test-bedding
Research and development	Decision making
Very cyber-secure	Research and development
High cost for the software (Free student version)	High cost for the software

Table 5 SmartWorldPro vs 3DEXPERIENCity

We can conclude from the table above that the SmartWorldPro offers a better solution for this project. Not only does it offer the same amount of functionality, but it can also be obtained free of charge by students. Nevertheless, 3DEXPERIENCity offers a lot of possibilities, but is still quite expensive to use.

Microsoft & Bentley Systems alliance VS Microsoft Azure Digital Twin

It is important to know the differences between Microsoft Azure and Bentley systems. Both programmes have their specific uses and properties. The following table describes the most important differences and their fields of application.

Azure	Bentley
Digital representation of real-world things, places, business processes, and people	Design, build, operate and maintain critical infrastructure
Can be applied for supply chain processes in order to have a real time view about the process	3D visualization of complex engineering data
Manufacturing goods	3D building information models (BIM)
Do not have to write the full code, it can be taken from CRM data and from azure Iot hub	Geographical information systems (GIS)
Free version for students	High cost for the software

Table 6 Azure vs Bentley

The table above gives an unambiguous conclusion on the differences between Microsoft Azure and Bentley Systems. Azure is a cloud computing software and is responsible for processing the most complex calculations and storing data. In contrast, Bentley Systems offers the possibility of further development. It provides access to applications that make 3D visualisations, city planners and ultimately Smart Cities possible.

5.6. Conclusion

The research above in the combination comparison table shows that Microsoft Azure, the cloud-based computing platform, is the best option. It will take care of the processing, storage and communication of the data. As the software focuses on an open source approach, it can be easily combined with other IoT based applications. Moreover, it can also be freely obtained as a student, which is a huge added value for this project.

In combination with Microsoft Azure, we recommend using Bentley systems or Smartworldpro from Cityzenith. Both systems are very powerful and give the possibility to process and create a full 3D digital twin model of the city. These systems are also freely available as students and therefore offer the best solution.

6. Information infrastructure of the Digital Twin cities

The information infrastructure of the Digital Twin of a city includes the following components:

- sensors that collect information;
- networks that transmit data;
- software systems that cleanse data;
- software systems that are engaged in data mining.

In order for DT to be able to provide synchronization between the state of the process in the real world and its virtual copy, we must provide it with the ability to receive, transmit and analyze the data flow from intelligent endpoints of IoT devices.

6.1. Sensors

Cities need reliable information to help both protect their citizens and to more efficiently manage their resources.

In the project Digital Twins I. Were selected different kinds of sensors who respond to different kinds of stimulus (such as heat, light, sound, pressure, magnetism, or a particular motion). These sensors send back information and this information is mandatory in our research.

In this chapter we will make a review of the sensors that we have to use in Digital Twins II.

The sensors listed below are crucial for the city to be able to carry out the necessary analysis. It is not possible to give a minimum value of the number of sensors. The number of sensors is directly proportional to the correctness and accuracy of the analysis.

Research shows that the city owns several sensors but does not make them public, so access is impossible. Making public data “open”is a crucial step for the further development of a smart city. Almost every city that develops, with the help of a digital twin, has public access to all public sensors. These applications will promote development.

We can reflect this in relation to Barcelona. The city is in a further development phase, where they are working on digitizing the city and making it more public. The current sensor information can be obtained from the public authorities of the local city website. This gives us the following cityscape, shown on the following map.

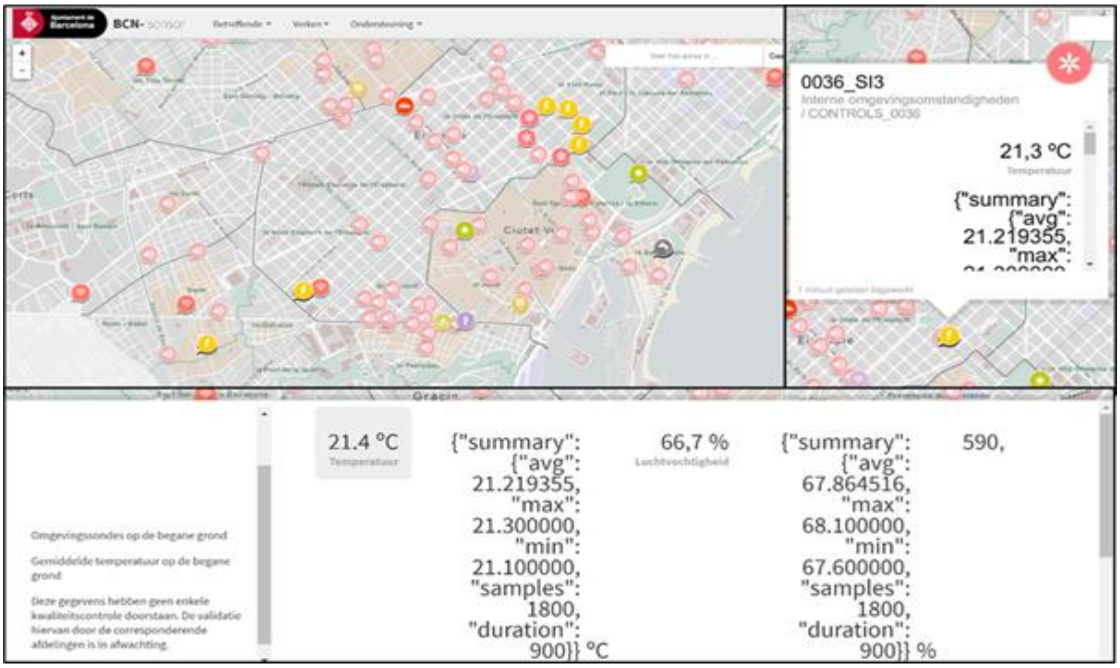


Figure 6 Sensors of Barcelona

As you can see, the city website gives a very structured picture of all the public sensors in the city. It gives enormous added value to these sensors and is an important step towards further development.

The sensors we found are listed below. The analyses were performed on the data obtained from Neapolis, this data was mainly outdated and not consistent enough to make definitive conclusions. To have unambiguous answers, one should obtain multiple measurement points and real time data.

Sensor	Sensor clasification	Sensor type	Data Neàpolis	present in city	access to real-time data
Atmospheric sensor			**X**	**X**	
	temperature sensor		x	x	
	wind sensor		x	x	
	UV sensor		x	x	
	solar radiation		x	x	
	pressure sensor		x	x	
Air quality sensor					
	Gas sensors				
	Elektromechanical sensor				
		infrared sensor (IR/NDIR)			
		Semiconductor sensor (Mox)			
		PID sensor (Photoionization)			
		Particle sensor			
Moisture sensor				**X**	
Traffic intensity sensor				**X**	
	Intrusive				
		embedded magnetometer			
		pneumatic tube sensor			
		inductive loop			
	Non-intrusive			x	
streetlight sensor					
container sensor					
	image-based sensing				
	ultrasonic sensing				
	infrared sensing				
sea level sensor					
	surface following tide gauges				
	Float based tide gauges				
	fixes sensor tide gauges				

Table 7 Sensors

Unfortunately, the table is not extended. This tells us that the publicly available data or the data generated in the city is not available or is not being generated. A further expansion of the available data is very important for further development.

6.2. Data cleaning

During the life cycle of sensors, the data produced by them can be "contaminated". Contamination of the incoming data manifests itself in the form of incorrect, abnormal or missing values. Contamination can be caused by human factors, sensor failure, communication line failures, planned sensor maintenance, etc. Contaminated data devalues its subsequent use in models based on data mining and neural networks. For use in the city's digital twin, the data needs to be cleaned up. Clearing data from sensors involves the following basic operations: searching for anomalies and restoring missing values.

Searching for anomalies involves finding intervals and corresponding values of the time series, which are significantly different from the values in all other intervals of this series. The restoration of missing values implies the generation of synthetic time series values instead of missing or erroneous displays based on a retrospective analysis of the time series values of a given sensor and/or sensor time series values that are geographically/logically close to this one. The solution of these problems can be performed both **by data mining** methods and **based on neural network models**.

When creating a digital twin of a city, an important component of the system is the data cleansing module, which ensures the search for anomalies and the restoration of missing values in the data streams coming from the system's sensors. The work of the data cleansing module is organized in accordance with the following basic principles. First, before entering data centers, the data stream of each sensor group must be processed in accordance with its own cleaning rules. Second, the cleaned data stream of each sensor group must be regularly monitored in accordance with its own verification rules.

The cleaning rule defines the area of correct values for each group of sensors (for example, the minimum and maximum readings, the dependence of the readings of the sensor of this group on the readings of the sensors of other groups, etc). The verification rule for a given group of sensors determines the frequency and method of assessing the probability of synthetically generated values instead of missing or abnormal ones. To assess the probability, various measures are used (for example, the normalized mean absolute error, the mean square error, etc.) and the limit values of these measures corresponding to the sensor group. The low chance of synthetic values revealed as a result of verification is a reason for revising the corresponding methods and algorithms implemented in the data cleansing module.

The next figure shows the data cleaning cycle, starting with the importing data into an algorithm, preprocessing and ending with the exporting already correct ones, with which it is possible to work further.

Figure 7 Data cleaning cycle

6.3. Foggy computing

To process information from multiple sources of the Internet of Things, it is advisable to apply the concept of fog computing.

Fog computing is a layered extension model of cloud computing that facilitates the deployment of distributed applications and services that take into account network latency, on the so-called fog nodes (physical or virtual) located between smart end devices and centralized (cloud) services.

Fog nodes are context sensitive and support a unified data management and communication system. They can be organized into clusters vertically (to support isolation), horizontally (to support service federations), or in relation to network proximity to endpoint smart devices. Fog computing minimizes the network response time of supported applications, and also provides end devices with local computing resources and, if necessary, network connectivity to centralized services.

While cloud computing provides potentially limitless resources for solving tasks (that require significant computing resources), fog computing provides:

- provision of computing resources in the immediate vicinity of the end devices;
- preprocessing data before sending it to the cloud;
- solving problems that require a minimum response time.

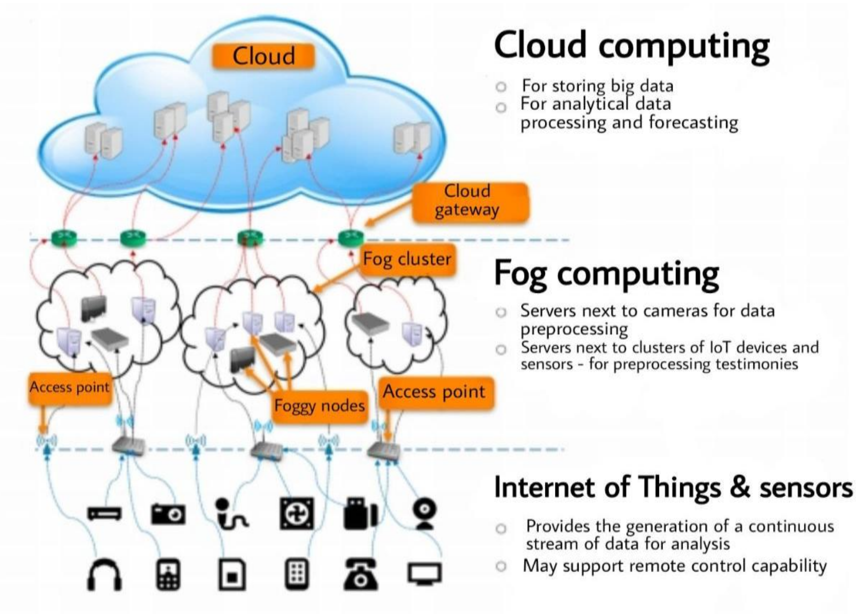


Figure 8 General model for processing data with fog computing concept

Figure 8 shows a general model for transmitting and processing data from sensors and IoT devices when implementing the fog computing concept. At the bottom of the model are IoT devices that generate data streams. These are data of different nature, different formats and different generation rates. IoT devices are directly connected to access points that provide their connection to the Internet.

In the immediate vicinity of the access points, an infrastructure of clusters of fog nodes is deployed, providing preprocessing of incoming data streams. According to the data, about 40% of the computing load for data processing in IoT systems will be provided on foggy nodes. Also, due to geographic proximity to IoT nodes, fog nodes provide low latency when two-way communication with end nodes is required.

The preprocessed data is transferred to the cloud for further processing. Unlike fog nodes, cloud computing infrastructure implies the provision of virtually limitless resources on demand to accomplish computational tasks.

An example of the use of fog computing to organize effective data processing is the task of preprocessing data from videcams. Today, the city can have hundreds of outdoor surveillance cameras, providing video monitoring services for a large area of objects of particular interest for urban management. If the resources of a single data processing center (cloud) are used for their analysis, then the task of transmitting a stream of video data over trunk channels from such a large array of cameras can be significantly complicated. To solve this problem, there are certain compromises that include:

- Strong compression of the video stream;
- Reduction of the resolution of the transmitted video information;
- Limiting the number of transmitted frames per second, etc.

Such decisions lead to the fact that the intelligent model used in the cloud system to analyze the incoming data may lack quality information to ensure effective operation. A possible solution to this problem is to deploy a network of fog nodes in close proximity to video sources. These nodes consume and perform preliminary analysis of video data in original quality, providing a solution to such problems of video data preprocessing as:

- Segmentation and selection of objects of interest;
- Background removal;
- Intelligent compression of the video stream.

This can significantly reduce the volume and quality of data transferred to the cloud, improving the quality of the data mining system.

7. Data

7.1. 3D/2D city map with corresponding sensors

The figure below shows a map with the current specifications and locations of the sensors.

Figure 9 Specifications and locations of the sensors

Observer	150590
Company	/
Online since	09/12/2018 12:23PM
Coordinates	41° 13' 40" N 1° 43' 31" E
Elevation	22.9m
Height	1.5m
Brand	Fine Offset
Model	WH Series

Table 8 Information Sensor 1

Figure 10 Sensor 1

Observer	FBC
Company	/
Online since	12/07/2020 2:40 PM
Coordinates	41° 13' 35" N 1° 43' 45" E
Elevation	10m
Height	5m
Brand	Bresser
Model	5-in-1

Table 9 Information Sensor 2

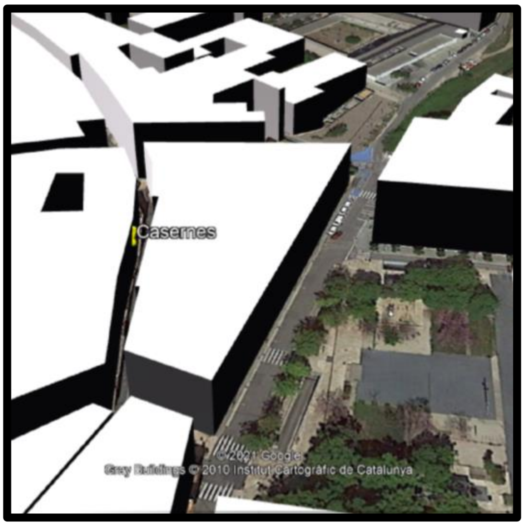


Figure 11 Sensor 2

Observer	Josep Miro - Garraf
Company	/
Online since	03/05/2020 5:13 PM
Coordinates	41° 13' 24" N 1° 44' 1" E
Elevation	14m
Height	5m
Brand	Davis Instruments
Model	Vantage Pro2 Plus

Table 10 Information Sensor 3

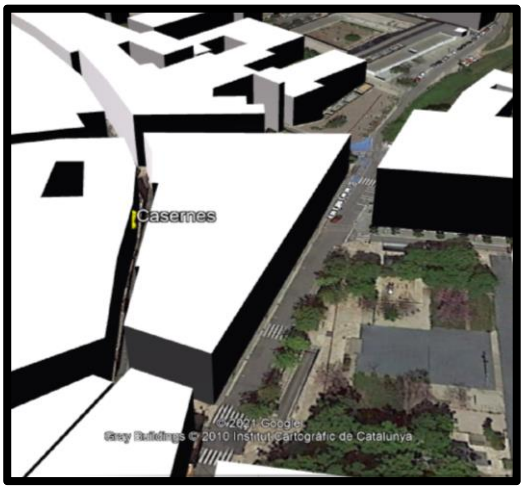


Figure 12 Sensor 3

The available public sensors are: observer 150590, FBC and Josep Miro - Garraf. They are the only sensors that provide a continuous data stream. Because there is no access to the original location of the sensors, they cannot be used for real time big data for the development of the digital twin.

7.2. Analysed data

7.2.1. Introduction and methodology

One of the needs to manage the Digital Twin from Vilanova i la Geltrú is to feed the algorithm with real world data, analysis and predictions of this one. With this information the algorithm will be able to compare the predictions and real world data and detect future problems or inform about the things that could be improved to have a better city in terms of pollution, resources, etc…

There is a lot of data that can be analyzed depending on the type of the sensors that the city of Vilanova has. The data we have collected has been provided to us by members of the city council in big data format in EXCEL. In the future this will not be necessary since the project will have direct access to all data warehouses. We have done some analysis from the big data that the sensors have produced. The methodology that we have followed to do the analysis is basically to get big data from the sensors that the city council has, once we have it we have different ways of working and interpreting the data. In our case we have used Microsoft Excel since it is one of the simplest tools to do a basic analysis. With this application we have differentiated the data in its different types, years, months, quantity, etc ... and from there we have made data tables to later make graphs to facilitate the interpretation of the data.[5][6]

TEMPERATURA MITJANA MENSUAL. **VILANOVA I LA GELTRÚ** **1994- 2004**

ANY	GENER	FEBRER	MARÇ	ABRIL	MAIG	JUNY	JULIOL	AGOST	SETEM.	OCTUB.	NOVEM.	DESEM.	ANY
1994	10,0	9,9	11,2	12,7	16,1	19,4	23,7	25,8	21,0	17,1	14,4	10,3	16,0
1995	10,3	11,9	11,3	13,1	16,6	19,3	24,1	24,2	19,4	18,5	13,2	11,0	16,1
1996	11,0	8,9	11,1	14,0	16,0	20,6	22,9	23,8	19,4	16,2	13,4	10,6	15,6
1997	10,1	11,3	12,3	14,1	17,7	20,7	22,4	24,1	21,4	19,4	13,9	10,7	16,5
1998	10,7	10,7	12,4	13,6	17,6	21,1	24,0	24,6	21,9	17,1	12,4	10,1	16,3
1999	9,7	9,9	12,2	13,9	18,1	20,7	23,5	25,2	22,1	17,9	11,2	10,5	16,2
2000	8,6	11,1	12,0	13,4	18,1	20,7	22,6	23,6	21,3	17,0	12,4	11,3	16,0
2001	10,6	10,2	14,2	13,8	17,7	21,7	23,6	25,2	21,0	20,1	12,4	7,9	16,5
2002	9,6	10,6	12,3	13,9	15,9	21,2	23,0	22,3	20,7	17,7	14,0	11,1	16,0
2003	8,9	8,4	11,4	13,6	17,5	23,9	25,3	26,3	20,8	18,2	14,0	10,5	16,6
2004	10,1	9,1	10,3	12,6	15,9	21,4	23,2	25,1	22,1	18,6	11,9	10,7	15,9
MITJANA	9,9	10,2	11,9	13,5	17,0	21,0	23,5	24,6	21,0	18,0	13,0	10,4	16,2
MÀXIMA	11,0	11,9	14,2	14,1	18,1	23,9	25,3	26,3	22,1	20,1	14,4	11,3	16,6
MÍNIMA	8,6	8,4	10,3	12,6	15,9	19,3	22,4	22,3	19,4	16,2	11,2	7,9	15,6

TEMPERATURA MITJANA DE LES MÀXIMES. **VILANOVA I LA GELTRÚ** **1994- 2004**

ANY	GENER	FEBRER	MARÇ	ABRIL	MAIG	JUNY	JULIOL	AGOST	SETEM.	OCTUB.	NOVEM.	DESEM.	ANY
1994	14,9	14,9	16,0	17,7	20,2	24,0	28,0	30,1	25,0	20,7	18,6	15,2	20,4
1995	15,4	16,9	16,2	17,9	21,4	23,5	28,8	28,6	24,1	22,6	18,0	15,2	20,7
1996	14,8	14,0	15,4	18,2	20,5	25,8	27,7	28,0	24,3	20,9	17,5	14,5	20,1
1997	13,7	16,6	17,8	18,8	22,6	25,1	27,2	28,9	26,1	24,4	18,5	15,0	21,2
1998	15,0	15,8	17,6	19,0	22,3	25,7	28,5	29,6	26,2	22,2	17,4	15,1	21,2
1999	14,6	14,8	17,2	19,3	22,8	24,8	27,8	29,2	27,1	22,4	16,1	15,3	21,0
2000	13,5	16,4	17,1	18,1	22,6	25,5	27,2	28,4	25,9	21,3	17,0	16,0	20,7
2001	15,2	15,5	19,4	18,9	22,7	26,3	28,3	29,6	25,5	24,9	16,9	12,5	21,3
2002	14,4	15,8	16,9	18,5	20,4	26,6	27,4	26,9	25,2	22,5	18,9	15,6	20,8
2003	14,2	12,3	16,6	19,0	22,0	28,6	29,8	31,6	25,4	20,8	17,7	14,4	21,0
2004	15,2	12,9	14,1	16,4	19,6	24,9	26,4	28,2	25,7	23,1	16,9	15,0	19,9
MITJANA	14,6	15,1	16,8	18,3	21,6	25,5	27,9	29,0	25,5	22,3	17,6	14,9	20,8
MÀXIMA	15,4	16,9	19,4	19,3	22,8	28,6	29,8	31,6	27,1	24,9	18,9	16,0	21,3
MÍNIMA	13,5	12,3	14,1	16,4	19,6	23,5	26,4	26,9	24,1	20,7	16,1	12,5	19,9

Figure 13 Example of the distribution of the temperature data in VNG in Excel

7.2.2. Meteorology analysis

Nowadays Vilanova i la Geltru has only three atmospheric public sensors which are localized in different places of the city; these sensors generate data about temperature, humidity, atmospheric pressure and precipitation.

Figure 14 Number of Atmospheric sensors

We have collected big data that was stored in the cloud of the Vilanova i la Geltrú city council, this big data corresponds to the meteorological data between 1994 and 2004. Using the previously described methodology in the previous section, we have carried out an analysis of data from the historical values of the data from these atmospheric sensors that will help the algorithm to make future decisions.

We have divided the data into information on temperature, humidity, precipitation and wind, in order to have more concrete and specific results.[7]

7.2.2.1. Temperature

Temperature is one of the most valuable data that we can get because of the important information that it provides. With the information that we find in the following graphs, the algorithm will be able to compare the real temperature at that moment with the historical temperature values and detect anomalies which can be beneficial (for example, a reduction in contamination implies an anomaly in temperatures of the city) or malicious that imply a problem for the city and the same advanced algorithm must be able to propose solutions to these detected problems.

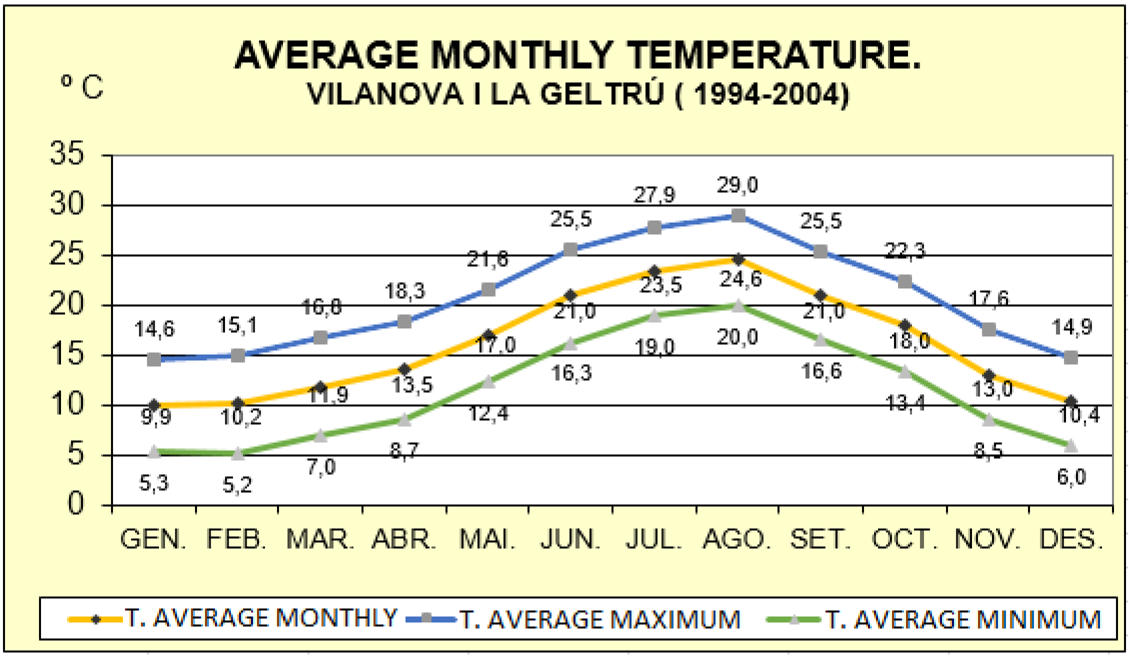


Figure 15 Average Monthly Temperature

In the town of Vilanova i la Geltrú the summertime is short, hot, humid and mostly clear. While the winters are long, cold and sometimes cloudy, there is dryness all year round. Temperatures generally vary between 10°C and 25°C throughout the year, and rarely drop below 5°C or exceed 29°C.

The warm season is from June to September with a monthly average temperature between 21°C and 24.6 °C. With an average maximum monthly temperature of 29 °C and an average minimum temperature of 20 °C , the warmest month of the year is August.

The cool season lasts 4 months, from November to March, with an average daily temperature of less than 13 °C. With an average temperature of 5 °C minimum and 14 °C maximum, January is considered the coldest month of the year.

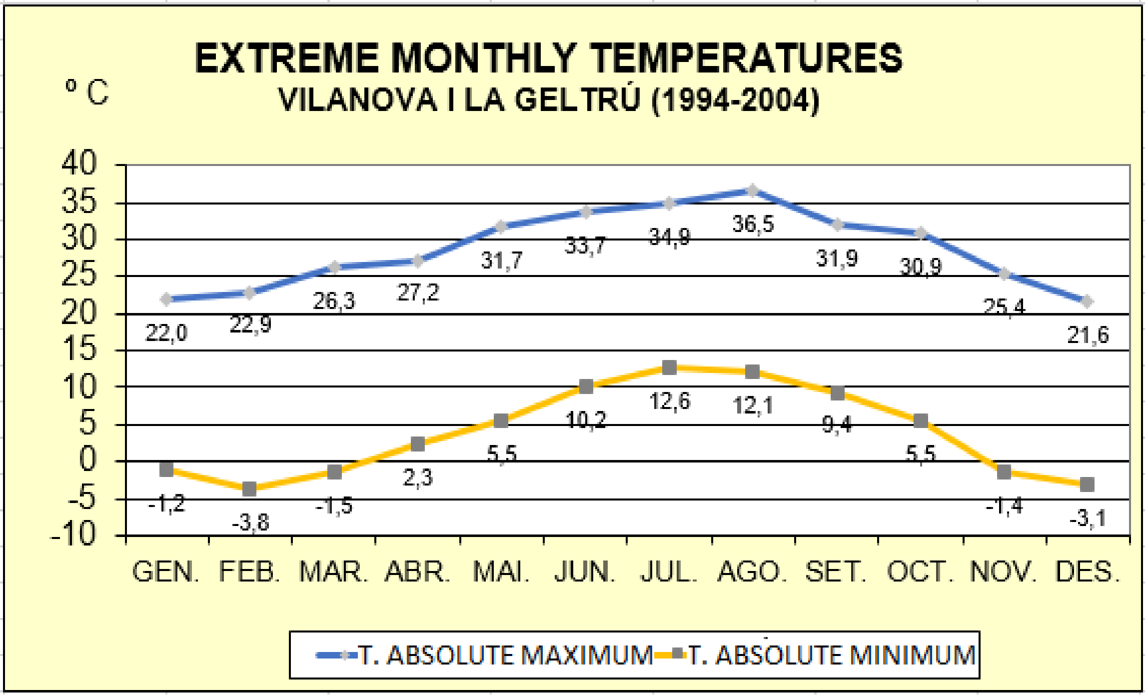


Figure 16 Extreme Monthly Temperatures

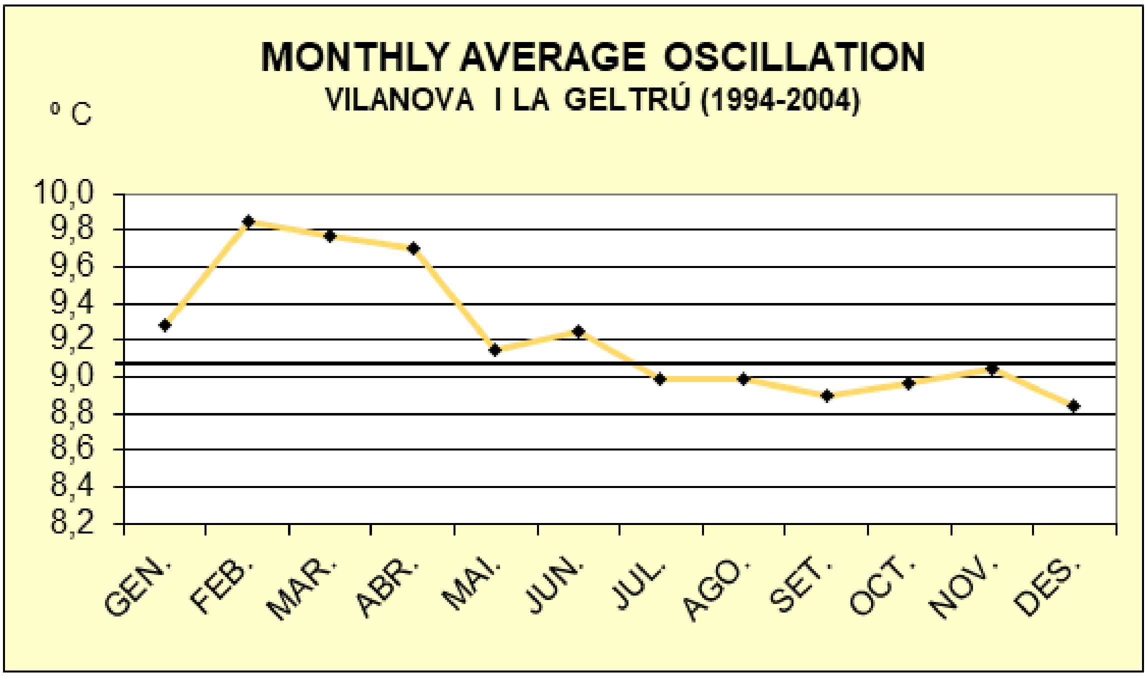


Figure 17 Monthly Average Oscillation

Vilanova i la Geltrú does not have a very extreme climate, the data show that the maximum number of degrees that has changed in one day is an average of 20 ° C. And during the year, no month suffers from large fluctuations in its temperature, except for February, which is perhaps the month with the most changes in its temperature.

However, on hot summer days in Vilanova i la Geltrú, special attention must be paid to watering, this should be done daily if possible, so that the water needs of the plants can be met. The algorithm that the Digital Twin controls, it is very important that it detects anomalies or reports the hottest days or the driest days so that the vegetation of the city is not affected and at the same time the irrigation resources are optimized.

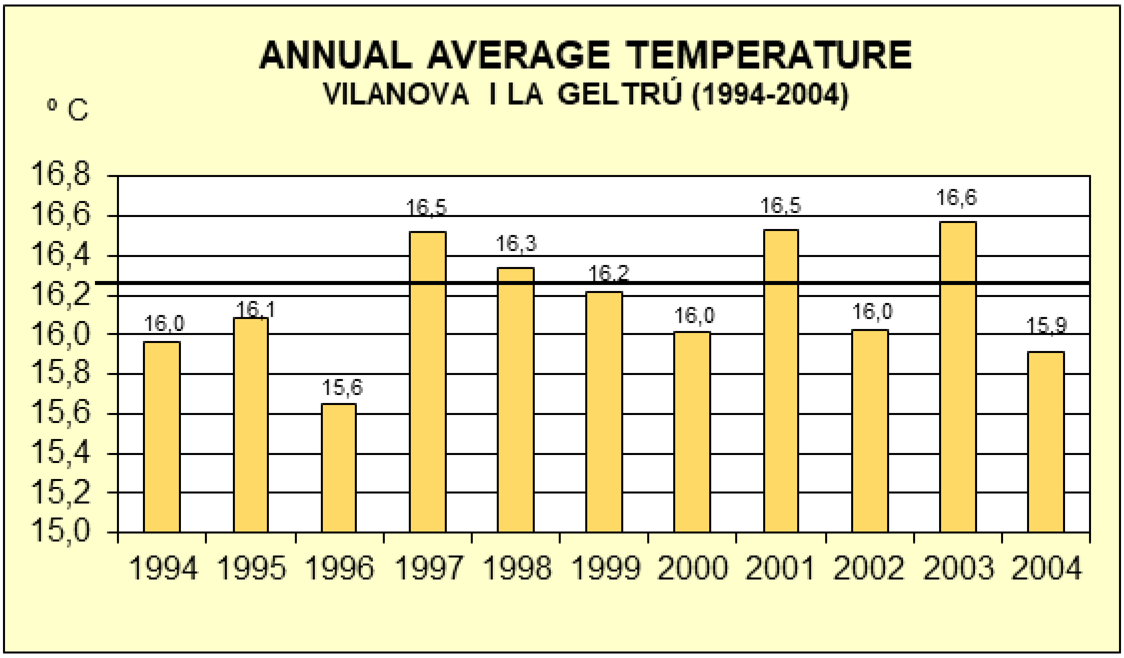


Figure 18 Annual Average Temperature

From 1994 to 2004, the city has maintained its Mediterranean climate with an average temperature of 16° C.

7.2.2.2. Humidity

Large gardens and parks in the Spanish city of Vilanova i la Geltrú with an automatic irrigation system are planned to be expanded that will save the laborious dragging of water, the complicated handling of the water pipeline, careful handling of the water, and plants as well as lawns will always have exactly the right amount of water (in accordance with the current weather data and rainfall). That is why having data on humidity in addition to temperature is very important for our Digital Twin to quantitatively reduce the resources used, and achieve an economic improvement in the city that could be used for other projects in addition to a reduction in pollution of the city.

Furthermore, humidity has an important role in the well-being of the inhabitants.

Recurring problems that high humidity can cause in people are: respiratory infections, asthma problems, allergies, worsening rheumatic diseases, bone pain, bad smells, discomfort in the home, defects in the ceilings and walls of the rooms , deterioration of furniture and other household equipment, etc.

This is why it is very important to take into account the real value of humidity at all times so that our algorithm can propose solutions when it detects high humidity values.

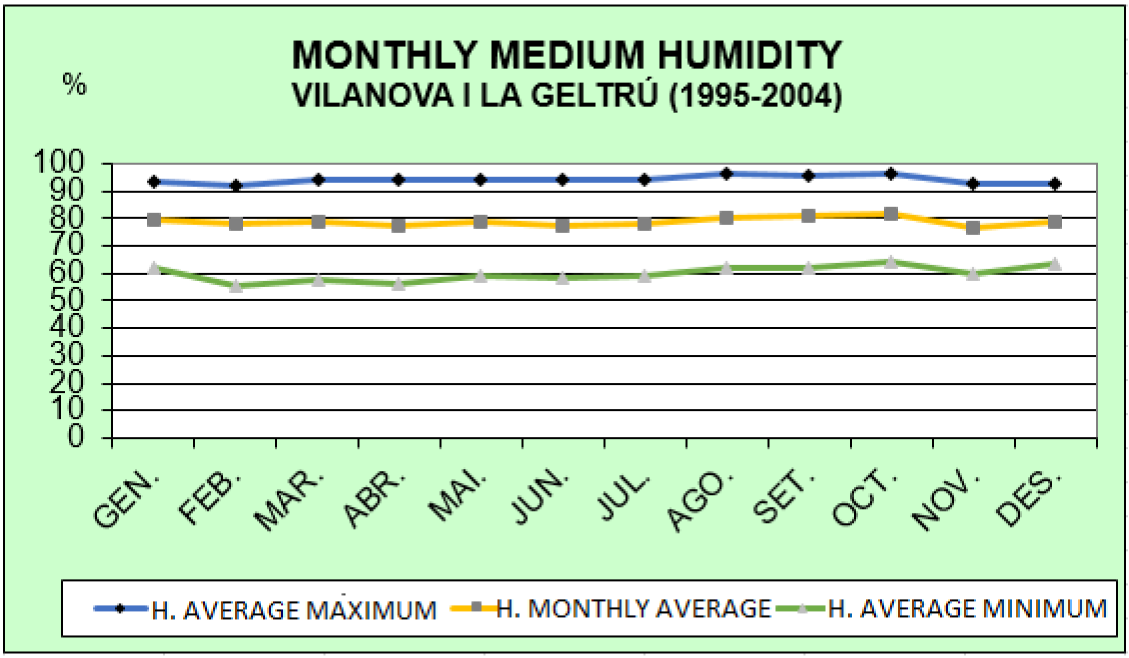


Figure 19 Monthly Medium Humidity

Throughout the year, the relative humidity is 80% on average, which means that of the total amount of water vapor that could be found in the air at the current temperature, it has 80%.

This is a very important piece of information that our algorithm has to take into account to optimize irrigation systems among other things.

Such a high percentage of humidity throughout the year is not beneficial to people's health because it is very important that residents maintain an adequate level of relative humidity in their homes to maintain thermal comfort and reduce the risk of occurrence. respiratory infections and allergies.

The algorithm cannot control humidity in the private properties of city dwellers, but it can control and optimize humidity in public buildings.

According to various studies, such as the one published in Environmental Health Perspectives, recommended indoor humidity values should be between 40% and 60% relative humidity. There are two reasons:

- The survival of airborne infectious bacteria and viruses (such as bronchiolitis, colds or flu) is reduced when exposed to relative humidity between 40 and 70%.
- Mite and fungal populations are minimized when the relative humidity is below 50% and they reach a maximum size at 80% relative humidity. And most fungi cannot grow below 60% relative humidity.

Therefore, most of the adverse health effects caused by humidity would be minimized by keeping indoor levels between 40-60%.

7.2.2.3. Precipitation

Precipitation is understood as a natural physical phenomenon that consists of the fall and arrival to the ground of water that is in the clouds, either in the form of snow, hail, drizzle or rain and this has an important influence on different labor fields. Measuring precipitation is a simple task that provides a host of truly valuable information, the increase in the concentration of carbon dioxide and other greenhouse gases are producing global climate change and in some places they represent more rains and in others severe droughts, This is modifying planning strategies over time with historical data as was traditionally done by specialists in agricultural climatology, who can change work strategies with crops such as planning the construction of larger lagoons or dams or even more wells for irrigation both in the present as in the future.

In addition, Vilanova i la geltrú, being a coastal city, has to take special account of the rainfall data since these can vary the sea level and cause anomalies in the port, the beach, the buildings, etc.

With the following data, the algorithm should be able to find anomalies in future rainfall and prevent the city from different problems related to agriculture, sea level, buildings, water supply, etc.

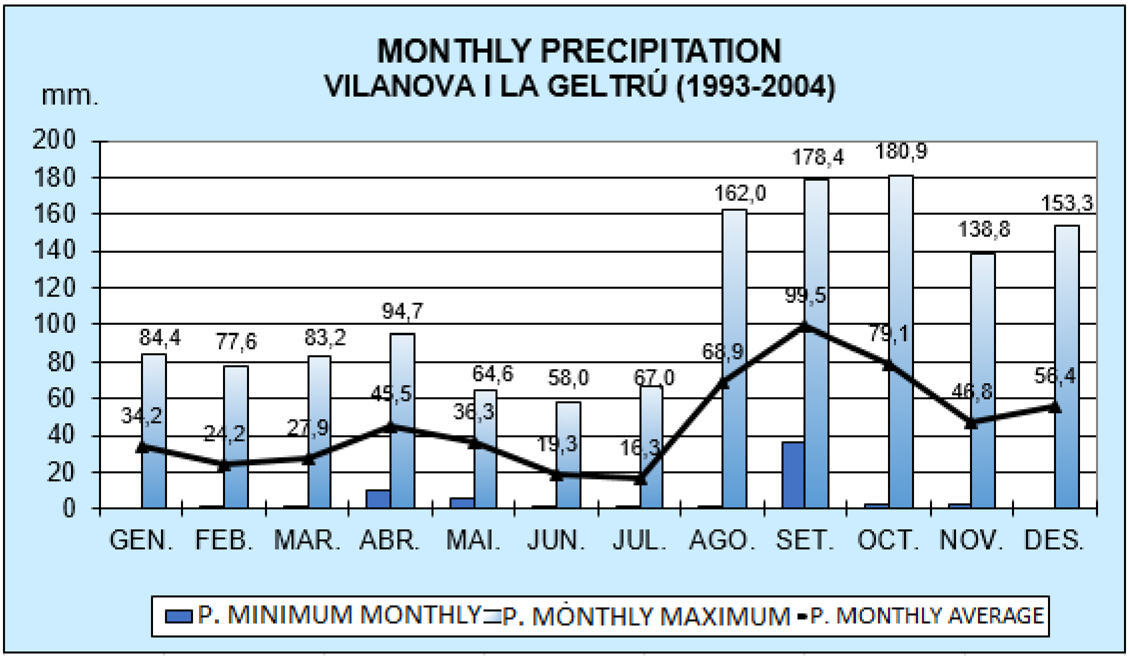


Figure 20 Monthly Precipitation

*Note: 1 millimeter of rainwater equals 1 L of water per m².

The city does not have stable rainfall during the year, between 1993 and 2004 it stands out that each month has different rainfall. The graph shows how September is the month with the highest rainfall on average, 99.5 liters of water per square meter, while July is the calmest regarding rainfall. As it does not have stable rainfall, the algorithm should be able to make corresponding decisions according to the month of the year.

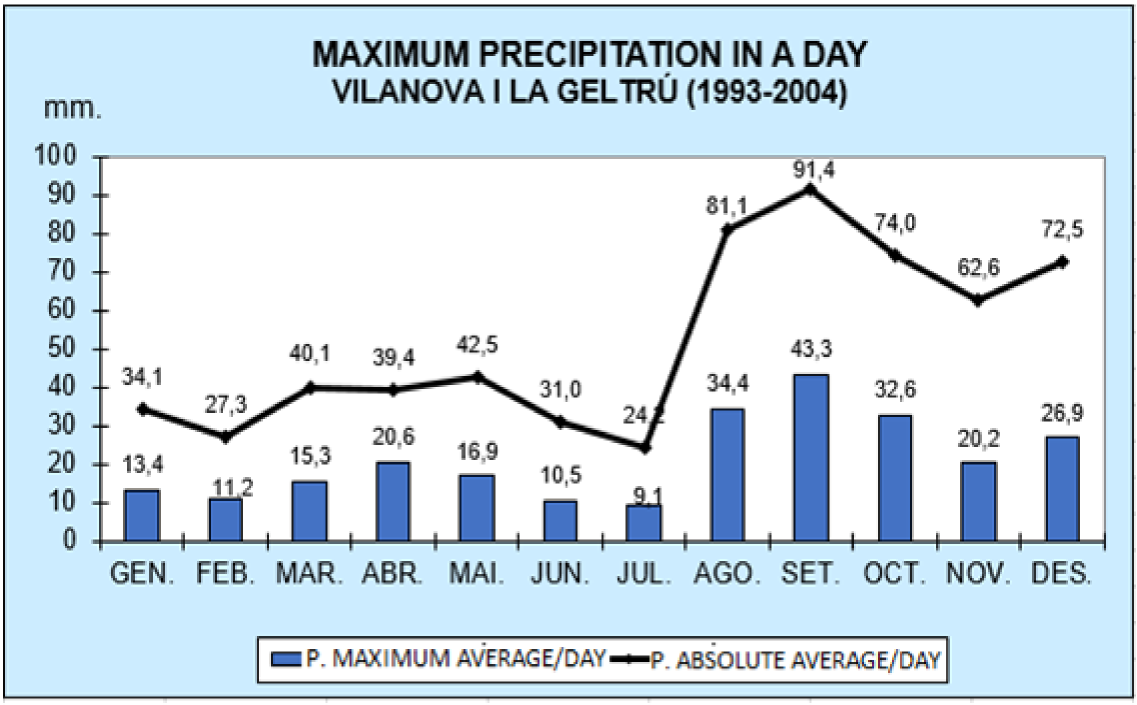


Figure 21 Maximum Precipitation in a Day

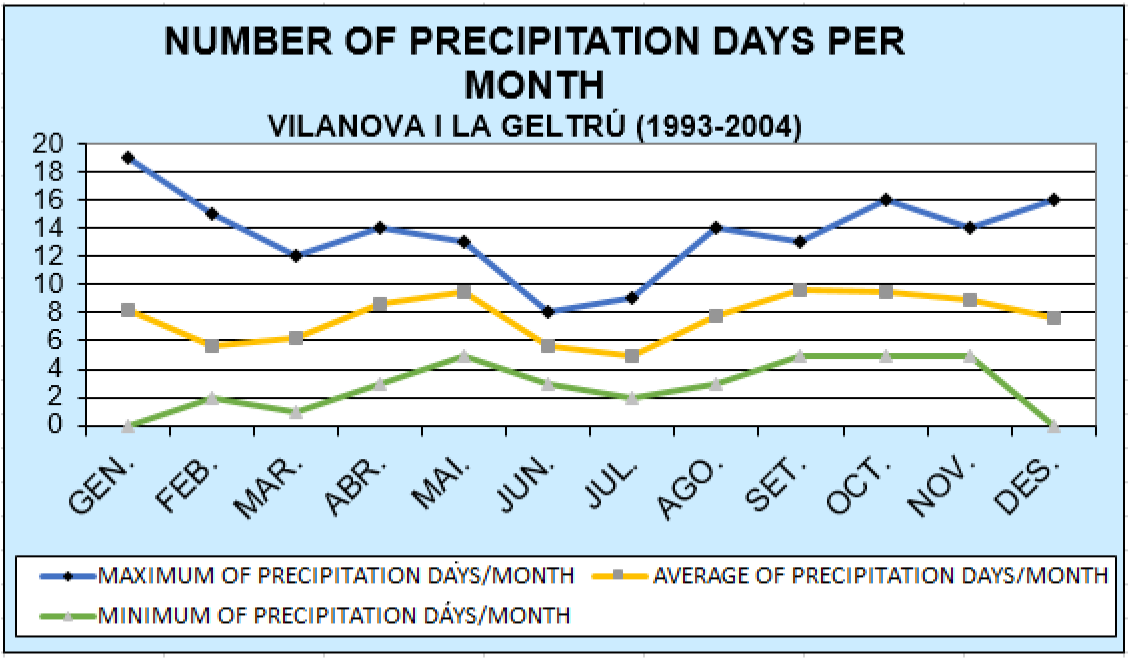


Figure 22 Number of Precipitation Days per Month

In these two graphs we can highlight that the number of monthly rainfall is stable throughout the year as it varies between 6 and 10 days of rain on average throughout the year, however the amount of rain varies throughout the year as we have commented.

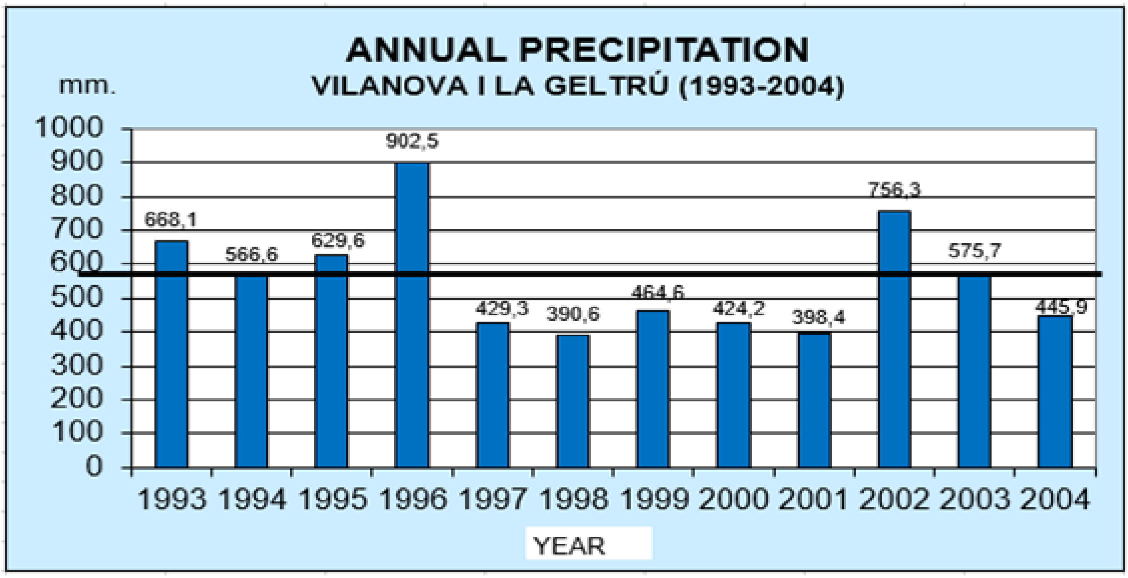


Figure 23 Annual Precipitation

The annual precipitation in Vilanova i la Geltrú during the years 1993 to 2004 is stable with an approximate average of 600 L of water per m² per year. Emphasizing that in 1996 there was an anomaly as there was twice as much rainfall as the following year (1997). This information is very important because the sensors could give anomalies in the data in the coming years due to the current climate change situation.

7.2.2.4. Wind

Wind is a very important piece of information that can be widely used by the city of Vilanova i la Geltrú. Its most frequent uses are:

- Installation projects, wind data related to different locations of the city can be useful for the development of installations.
- Building or construction projects, on the one hand, the wind is related to the structural design. Thus, design and calculate structures, considering (in addition to many other parameters) wind conditions (especially maximum wind speed and wind direction). It is true that in Spain there is a regulation (CTE: Technical Building Code), with considerations on the wind related to the design and calculation of structures, depending on the geographical location. But knowing the specifics of VNG can help you get more accurate results. On the other hand, it is also related to the natural ventilation of buildings.
- Useful for pollution studies and estimations. Indeed, depending of wind data can be estimated the transmission and propagation of a polluting source (for example, the chimney of an industry, etc.).

- Useful for agricultural and forest purposes.
- Useful for forest fire management (firemen)
- Useful for nautical sports (sailing, windsurfing, kayak, etc.).
- Useful for other kind of sports (paragliding, paramotor, etc.).

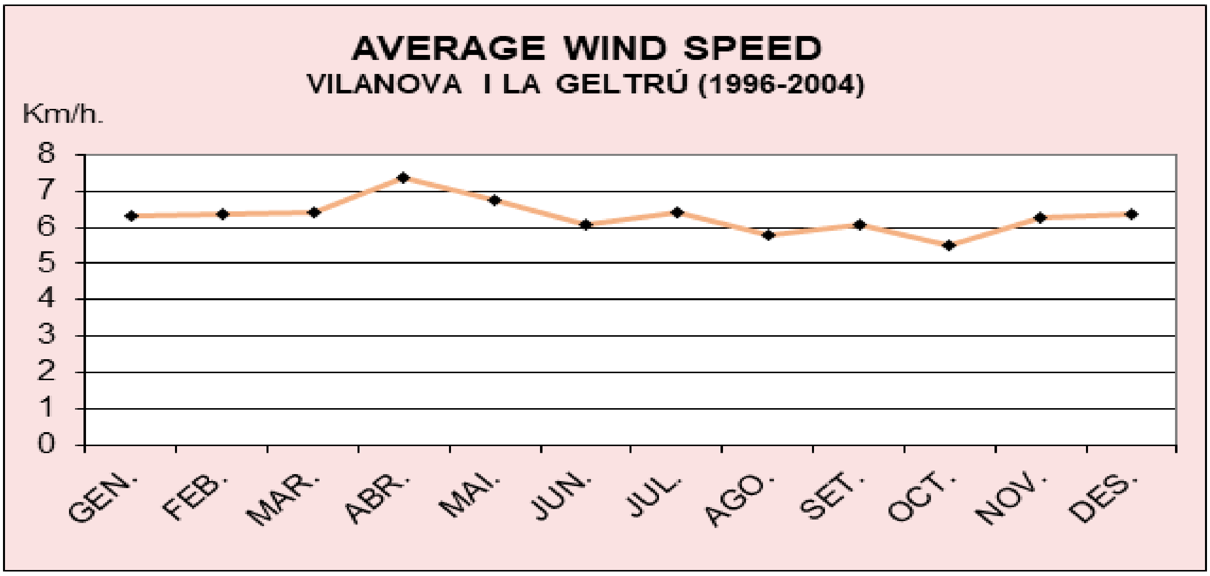


Figure 24 Average Wind Speed

Vilanova i la Geltrú has an average wind speed that according to the Beaufort scale (empirical measure of the intensity of the wind that is based mainly on the state of the sea, the waves and the strength of the wind) corresponds to level 2 due to a speed of the Average wind of approximately 6.2 Km / h, this means that the air can be perceived on the skin but that it is very soft, ideal for agriculture and industry.

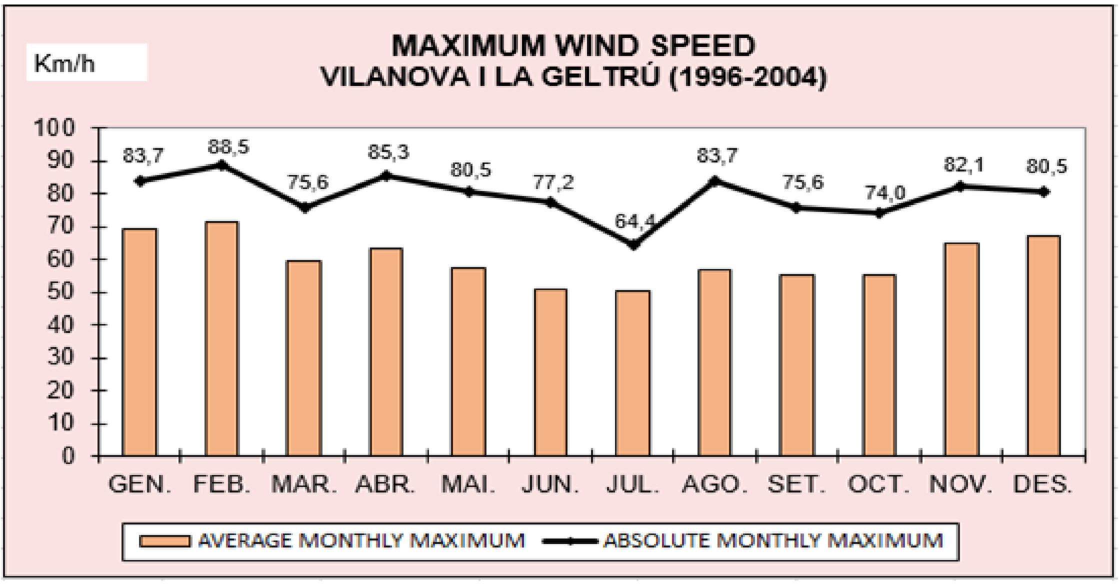


Figure 25 Maximum Wind Speed

The maximum wind speed during the different months of the year is approximately the same 65 km / h, which corresponds to level 8 in the escale Beaufort, which is very high since it can seriously damage trees and even buildings and is considered a very high wind.

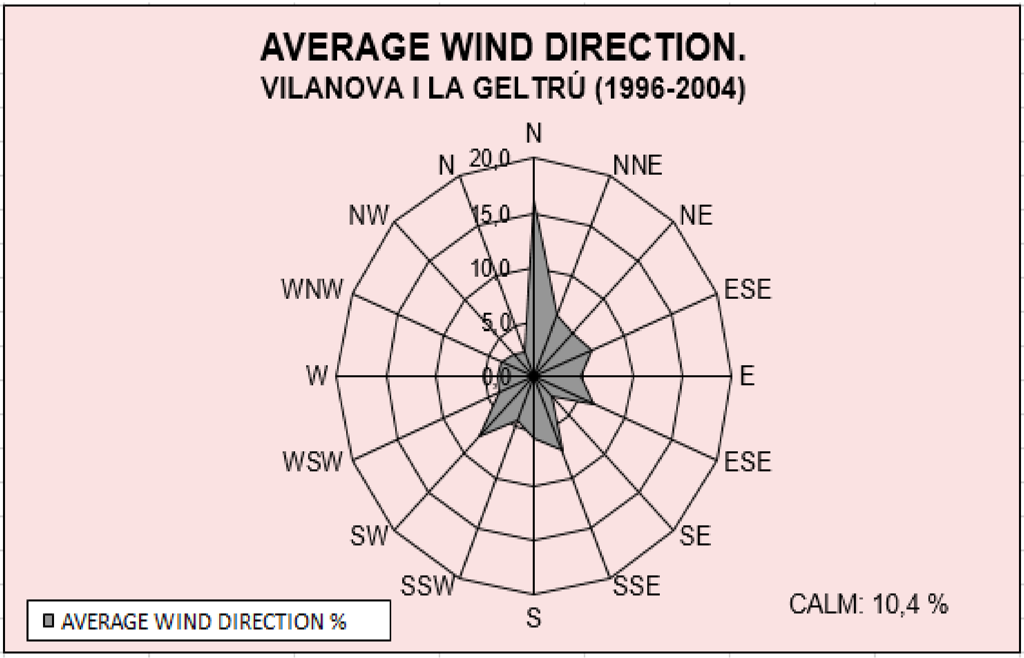


Figure 26 Average Wind Direction

This graph shows the mean wind direction during the year, where north stands out as the main wind direction between 1994 and 2004.

7.3. Example of Big Data

Digital Twin II has received different types of data collected by the sensors in order for the project to carry out a theoretical analysis of it. In this section we will show different examples of big data and its possible use.

7.3.1. Example of environment Big Data

The following figure corresponds to a Big Data file granted by the Vilanova i la Geltrú city council that corresponds to the environmental sensors during confinement.

codi_eoi	nom_estacio	data	magnitud	contaminant	unitats	tipus_estaci	area_urbana	codi_ine	municipi	codi_comarc	nom_comarc	h01	h02
8307012	Vilanova i la	2020-07-27T(	12	NOX	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	9	7
8307012	Vilanova i la	2020-07-28T(	12	NOX	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	16	11
8307012	Vilanova i la	2020-07-28T(	14	O3	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	58	57
8307012	Vilanova i la	2020-07-27T(	7	NO	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	1	1
8307012	Vilanova i la	2020-07-28T(	7	NO	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	1	1
8307012	Vilanova i la	2020-07-28T(	8	NO2	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	15	11
8307012	Vilanova i la	2020-07-29T(	12	NOX	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	17	16
8307012	Vilanova i la	2020-07-30T(	12	NOX	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	12	14
8307012	Vilanova i la	2020-07-30T(	1	SO2	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	1	1
8307012	Vilanova i la	2020-07-29T(	7	NO	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	1	1
8307012	Vilanova i la	2020-07-30T(	7	NO	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	1	1
8307012	Vilanova i la	2020-07-29T(	8	NO2	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	16	14
8307012	Vilanova i la	2020-08-03T(	6	CO	mg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	0.2	0.2
8307012	Vilanova i la	2020-08-01T(	12	NOX	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	23	15
8307012	Vilanova i la	2020-08-02T(	12	NOX	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	23	14
8307012	Vilanova i la	2020-07-31T(	14	O3	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	35	39
8307012	Vilanova i la	2020-07-31T(	1	SO2	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	1	1
8307012	Vilanova i la	2020-08-02T(	1	SO2	Âµg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	1	1
8307012	Vilanova i la	2020-07-31T(	6	CO	mg/m3	traffic	suburban	8307	Vilanova i la	17	Garraf	0.3	0.2

Figure 27 Example of Environment Big Data in Excel

The image shows an excel file where it has been autoredated with the data generated by the environmental sensors. There are different points to highlight, in this case the data shown are the contamination levels of one of the days of the alarm state where the contamination statistics were reduced due to the lockdown. In the excel there are different relevant parameters, on the one hand there is the column of polluting gases where the different gases analyzed by the pollution sensor are shown and on the other hand the columns "h01" and "h02" show the magnitude of these gases at one in the morning and two in the morning respectively.

The use of an algorithm for the project using Digital Twins II in this case, has the objective of being able to read this type of files full of Big Data ordered and collected by the sensors without the human factor having to intervene. The algorithm will decide if the data it has read from the sensors is important at that moment or if it detects any anomaly and through machine learning and possible future programmed solutions it will perform a certain function.

In this case, due to the state of alarm, the contamination levels dropped and the algorithm will detect it as a positive anomaly.

7.3.2. Example of water supply Big Data

In the following figure you can see an example of Big Data related to water supply, this information has been provided to the project by the city council and the city water company.

Data: 01/05/21 Hora: 0:01

Estacio	Comptador	Cabal min. m3/h	Cabal max. m3/h	Acumulat m3	Volum m3 30/04/21
Diposit n.1	Entrada ATLL	0.00	520.00	802512	5460
Diposit n.1	Entrada arqueta Coladors	3.00	15.00	808221	170
Diposit n.1	Sortida a xarxa 600	73.00	318.00	780256	4972
Diposit n.1	Sortida a Collada-Aragai	10.00	31.00	75789	489
Diposit n.1	Impulsio 50 a Collada-Alta	4.24	20.10	36909	245
Diposit n.1	Impulsio a Dip.Camping	0.00	115.00	786150	7
Diposit n.1	Sortida a Coladors bomb.a	13.30	29.30	58114	477
Repartiment Mina	Mina Coladors (Mina+Pou	3.00	15.00	808221	170
Repartiment Mina	Sorea Sitges	49.00	51.00	566070	1200
Repartiment Mina	Camping auxiliar	0.62	1.53	196135	17
Repartiment Mina	A Diposit 1(Mina-Sorea+P	3.00	15.00	808221	170
Repartiment Mina	Pou Coladors	0.00	26.70	135072	2
Repartiment Mina	Mina	53.00	64.00	1374291	1370
Diposit n.2	Sortida a Sector Moli de V	2.50	13.80	998989	142
Diposit n.2	Sortida a xarxa	0.00	0.00	12639	0
Diposit n.3	Entrada de Dip.Costa-ATLL	0.00	1120.00	985710	6810
Diposit n.3	Sortida a xarxa	139.82	388.68	27422610	6582
Diposit n.3	Subministrament poligon	0.00	21.00	51890	320
Diposit n.3	Bombeig Prysmian	0.00	19.50	11469	95
Diposit n.3	Impulsio a Torre Veguer	0.00	0.35	398	2

Figure 28 Example of Water Supply Big Data in Excel

In the excel you can see in the first two columns the type of station or tank and the water meter respectively. A tank has different water meters because it has different outlets and it is very important to know the flow information that passes through each of the inlets and outlets to detect possible anomalies. And then, we have different columns that determine the maximum or minimum flow that has passed through each water meter, the accumulated and the available volume.

For the algorithm, this information is very useful because in the future when it reads the Big Data generated by the water supply sensors, it will be able to optimize some of the supply areas so that the water reaches all the needs of the inhabitants of the city and there are no problems of water shortage or flooding.

7.3.3. Example of bad Big Data

Previously we have seen two examples of Big Data in an excel file generated by different sensors. In these examples it has been possible to observe how the data followed an order and a kind of pattern and information could be extracted from the different rows and columns easily. Next, we are going to see an example of bad big data created by environment sensors.

name	time	dada	nom	sensor					
sensor	2020-10-27T1	1015.66	medi	ambient	ext	1	pressio		
sensor	2020-10-27T1	3.017	medi	ambient	ext	1	voltatge	condensadoı	2
sensor	2020-10-27T1	609	medi	ambient	ext	1	concentracic	CO2	
sensor	2020-10-27T1	0	medi	ambient	ext	1	estat	sensor	CO2
sensor	2020-10-27T1	###########	medi	ambient	ext	1	temperatura		
sensor	2020-10-27T1	603	medi	ambient	ext	1	concentracic	CO2	LPF
sensor	2020-10-27T1	37.736	medi	ambient	ext	1	raw	ir	
sensor	2020-10-27T1	###########	medi	ambient	ext	1	humitat		
sensor	2020-10-27T1	23.11	medi	ambient	ext	1	temperatura	sensor	CO2
sensor	2020-10-27T1	37.767	medi	ambient	ext	1	raw	ir	LPF
sensor	2020-10-27T1	22.96	medi	ambient	ext	1	temperatura	barometre	
sensor	2020-10-27T1	3.052	medi	ambient	ext	1	voltatge	condensadoı	1
sensor	2020-10-27T1	2.755	medi	ambient	ext	1	bateria		
sensor	2020-10-27T1	2.932	medi	ambient	int	1	bateria		
sensor	2020-10-27T1	###########	medi	ambient	int	1	temperatura		
sensor	2020-10-27T1	###########	medi	ambient	int	1	humitat		
sensor	2020-10-27T1	1016.02	medi	ambient	int	1	pressio		
sensor	2020-10-27T1	0	medi	ambient	int	1	llum	ambient	visible
sensor	2020-10-27T1	0	medi	ambient	int	1	llum	ambient	ir
sensor	2020-10-27T1	0	medi	ambient	int	1	lluminancia		
sensor	2020-10-27T1	667	medi	ambient	int	1	concentracic	CO2	

Figure 29 Example of Bad Big Data of Environment

The figure shown is a capture of a part of the Excel data. It can be seen that there is no data that can be extracted or predicted, there is simply disordered information that does not follow any pattern and that cannot be analyzed.

The objective of showing this example of bad Big Data is to raise awareness of the importance of the order of the data generated by the sensors, since if the data does not follow any order or a pattern, the big data becomes information that cannot be used. and data collection is useless.

In the future of the project, special emphasis will be placed on the correct ordering of Big Data so that the algorithm can read it without difficulty. Big Data in itself is useless if it is not in order.

7.4. Conclusions

In the future of the project, the data analysis will be carried out using the same algorithm that will take care of all the functions, receive the Big Data, read it, compare it and propose solutions if it deems it appropriate.

The next step in the process after doing a data analysis is to propose solutions for possible improvements in the future. In our case, the data that the city council provided us was very scarce, which is why after doing the data analysis the solution proposal for this analysis was unnecessary because we did not have enough information. In the future of this project, the solution proposal will be very important since the algorithm needs this information to carry out its work.[8]

8. Big Data Management

8.1. Private data vs public data

It is extremely important to know the difference between public and private data and the use of it. Misuse of these data can lead to legal penalties and fines. The fine depends on the order of violation. They can range from 2% to 4% of the firm's worldwide annual revenue from the preceding financial year. Public data is maintained controlled and anonymized by the National Statistics.

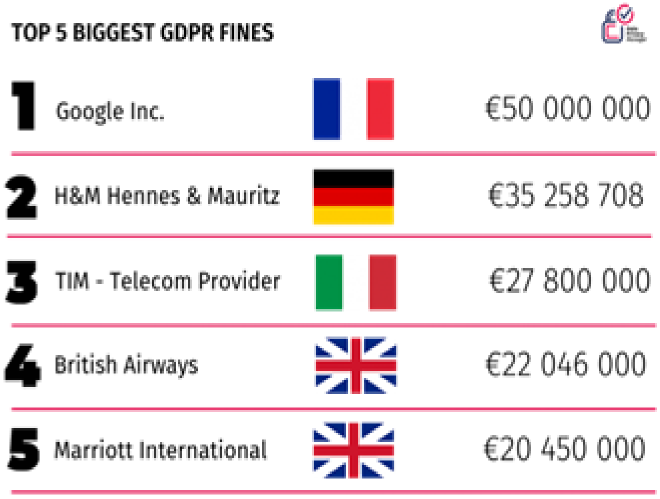


Figure 30 Top 5 Biggest GDPR Fines

To give a clear idea of what this 2-4% penalty means, you will find the fines for the 2021 above. These are huge fines, which are a deterrent even for large corporations. It provides a huge motivation, even for the biggest companies to follow the regulations.

Offices, spread across Europe and the U.S. The two main characteristics of public personal data are its reliability and stability. We are talking about anonymous stable data that is applied to 100% of the population. This can range from family compositions to education levels.

When we talk about publicly generated data such as traffic data, it is more complex. Some data will not be completely independent of the public, but its use is still considered public. One will have to be very careful with the use of it and take into account the legislation under the GDPR.

Eva Blum-Dumontet, research officer at the NGO Privacy International, underlines the fact that “people should always know that their data is being collected, and that these can be accessed and deleted”. Moreover, a city’s services should not be accessible only to citizens willing to surrender their data: “All the initiatives developed by a smart city should be carried out in the name of public interest.

and not in the one of companies providing cities with the technologic infrastructure.” (commission, 2018)

Private data is going to be data that is personal. It is data that is both unreliable and unstable and comes primarily from bulk or third-party entities.

The data is fully regulated under the GDPR and must be strictly monitored. However, this data is less useful for the development of our project. The data needs strict tracking and cleansing before it is functional, so for a start-up smart city like Vilanova (which is still in its start-up phase) it is not so applicable.

This background information forms the basis of the general legislation on data use. For further information on the use and processing of data, please refer to Regulation (EU) 2016/679 (General Data Protection Regulation).

8.2. Pros/cons of using Big Data Management

When we talk about the concept of big data management, we should first get the concept of big data. By looking at applications such as digital twins of cities, we see that data is being generated all around us. (IDC, 2020) More than 59 zettabytes (ZB) of data will be created, captured, copied, and consumed in the world this year, according to a new update to the Global DataSphere from International Data Corporation (IDC). To give us a better idea about the size of this data, Every individual as early as 2020 but generate 5500 Gigabytes (GB).

The description of big data does not stop here. Many definitions have been formed around big data but the most widely used, is that of "data warehousing pioneer" Douglas Laney. He defines big data as previously described under several V's, namely: volume, velocity and variety.

This huge amount of data is an important factor in the further development of many applications. These can range from the prevention, diagnosis and repair in healthcare, to targeted and specialized marketing or even to the development of digital twins of cities.

All this data, does not come alone with its benefits. Below are the pros and cons of bigdata management.

Pro’s	Con’s
Quick detection of anomalies like errors or frauds	Real-time analysis requires constant data collection. Major changes to current periodic collection strategies. Big investments.
Defense mechanism against proprietary information	Real-time analysis requires advanced software’s, algorithms and data scientists.
More direct and effective strategies, by predicting and experiment with the datasets Hereby reducing costs and improving productivity.	Mishandling or misrepresenting data can cause serious legislation because of the GDPR (European Union’s General Data Protection Regulation)

Table 11 Pro/Cons of Big Data management

Savings, because of more & effective strategies: As mentioned, data is all around us and creates. However, little or none of it is used. This unused data is also called "dark data". In a study by IMB Andrew Trice described the following.

Consider these statistics:

- 2.5 Quintillion bytes of data created every day. (That's 1,000,000,000,000,000,000,000,000 bytes)
- 90% of the data in the world today has been created in the last two years alone.
- Every minute 1.7 MB of data is created for every person on the planet. All 7.3 billion of us.

Unstructured data - "dark data" - accounts for 80% of all data generated today.”

The world's population will only increase, by 2050 it will be 2.5 billion people. This will especially affect the densely populated areas. In order to meet the goals of the Paris Agreement, we are obliged to take action. Big Data management is currently one of the main ideas to obtain more energy efficient applications.

Another study by the multinational ABB shows that world electricity consumption will increase by 60% in the next 20 years and almost 60% of this will be used for domestic purposes.

The concept of SmartCities is one of the best applications to make this energy consumption as efficient as possible. According to the same study, the Smart city market is expected to increase by 20% in the next 5 years. This huge increase is due to the awareness of the major climate issues, combined with the huge population increase.

The Smartcities can monitor the energy consumption inside buildings and use energy management systems.

By using these connected devices, the city can reduce its energy consumption and operating costs by up to 30%. This is due to the fact that consumption can be monitored, rapid action can be taken in the event of errors, and well-founded and predictive action can be taken.[9]

8.3. Risks of problems appearing from Big Data Management

8.3.1. Data regulation

Data regulation is a very actual point. It is divided into open data regulation and private data regulation. In recent years the European Union has made a reputation of being data protection champions. They're trying to achieve data protection and open data at the same time. When making open data accessible for everyone and by implementing GDPR "general data protection regulation" there regulates the private data by 6 categories. The GDPR explains in "Retical 40", in order for processing to be lawful, personal data should be processed on the basis of consent of the data subject concerned or some other legitimate basis. In other words, the data used should meet the terms in one of these categories. The GDPR explains these legal bases: (Union) [10]

1. [1] Processing shall be lawful only if and to the extent that at least one of the following applies:

   (a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;

   (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;

   (c) processing is necessary for compliance with a legal obligation to which the controller is subject;

   (d) processing is necessary in order to protect the vital interests of the data subject or of another natural person;

   (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;

   (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

Figure 31 Citation GDPR

Compare this with the data needed for our project, the used data will mostly come from public organisations. Thus the information will be open data and open for everyone to use. As the European Union decided, data generated with public money should be freely available for reuse by the public. Here for the used data for our projects mostly come from public bodies and should be open, free for the public to use.

The data will have fewer regulations. For more information about regulations considering data use, we direct you to the local authorities and local regulations from the country where to project is in use.

9. Data Security

9.1. Cyber security introduction

The set of new technological solutions that allow organizations to better manage their information, commonly known as "Big Data", have a growing role in all types of public and private organizations. Today more data is generated, more information is analyzed and more analysis results are consumed than ever before. The time that elapses between the appearance of the data and the decision-making based on it is less and less, and the economic and social importance of these decisions is increasing.

Big Data is characterized by the real-time collection, storage and analysis of huge amounts of data in different formats. Organizations are putting data at the center of their management and operation. This broadens their possibilities of action and competence but brings with it new risks, possibly not fully contemplated in conventional security scans.

By its very nature, Big Data gains insight from where there is only data at first. A potential conflict of ownership and privacy appears here. Those who generate the data are not always aware of the value they contain, even economic, and they transfer them to companies and public administrations without reflection or compensation. Unexpected situations arise for the generation of personally identifiable information, so that the metadata exceeds in importance to the data itself. Voices are raised that warn of new forms of social control and that demand the free distribution of knowledge of human dynamics that Big Data allows.

Although still in the initial stages, it is not an exaggeration to say that fields of application such as "Smart Cities", "Internet of Things" and the plethora of achievements of "Machine Learning" are going to change our societies and that along with their undoubted benefits will come risks . They will join those that we already accumulate in the energy distribution, telecommunications or production of basic goods networks. They could be even more vulnerable: Malicious data injection attacks on Big Data-based city or factory monitoring and control systems are conceivable, for example.

In a Smart City, all interconnected cyber-physical devices and processes generate huge amounts of data, much of it in real time and on a very granular scale. The collection, processing, transfer and use of data enables smart living, an instant connection with / between all citizens, and creates the possibility for cities to function more efficiently, productively, sustainably, fairly and transparently. But, on the other side of the story, several problems occur in the huge data machinery that is a smart city: internal and external parts cannot be trusted, new threats affecting confidentiality, integrity, accessibility, data protection and privacy is continually being pointed out, smart city technologies are still in their imagination, there are no standards of use and many technical difficulties need to be overcome.

This section wants to report on the dangers that could exist in the future smart City created from the Twins digital concept in which Big Data plays a very important role.[11]

9.2. Data vulnerabilities in a Smart City

In a smart city, objects are connected to provide seamless communication and contextual services. A wide variety of elements are used in a smart city. Some of them are very sophisticated embedded systems, such as smartphones and televisions, tablets, printers, medical devices, SCADA (supervisory control and data acquisition). In addition, many sensors are used to monitor air quality and pollution. car and pedestrian traffic, resistance of bridges and road infrastructure in general, crime rates and surveillance, energy and water consumption, waste management, etc., with the aim of forming a perception and recognition layer to collect data and identify the physical world.

Smart things suffer from hardware limitations (power and computational restriction, memory restriction, tamper-proof packaging), software restrictions (embedded software restriction, dynamic security patch), strict network requirements (mobility, scalability, multiplicity of devices, multiplicity of communication media, multiprotocol networks, dynamic network topology). These resource limitations restrict the inclusion of suitable security mechanisms, such as cryptography, directly in smart objects. Consequently, the designers put security aside, hoping it could be added later, and resistance to attacks is generally losing the race to other design factors such as good performance, small form factor, and low power consumption. A study by Hewlett-Pack-ard showed that 80% of things in IoT do not require passwords of sufficient complexity and length, 70% allow an attacker to identify valid user accounts through account enumeration, 70% % use unencrypted network services and 60% have security concerns with their user interfaces.

Data collected by smart things is at the heart of smart cities. The problem is that it is sensitive data, often collected without our explicit consent. For example, messages, personal photos, appointments, bank account information, contacts and others are stored on our smartphones with full conscience, with more or less security measures in place. From another range of devices, the thermostat reports their location (including zip code), temperature data, humidity and ambient light data, time and duration of activation; These data can be used to determine the domestic habits of a citizen; Medical bracelets store heartbeat and sleep patterns, collecting biometric and medical data that reveal people's physiological status. It is obvious that if this valuable data is not treated well, significant privacy problems can occur.

Also, most things in a smart city are not personal and are neglected. Their physical security is not guaranteed, especially in public networks, control of objects can be lost and cascading failures can appear, caused by the interconnectivity of a large number of devices, difficult to protect simultaneously. Some connected things and their firmware are protected by trade secrets. Furthermore, the legal framework is not yet appropriate and the legal responsibilities are not clear enough.[12]

9.2.1. Smart spaces

A smart space is described as a collection of smart things and other relatively powerful computers / gateways that manage and serve them; a fusion of physical and digital spaces, which have some kind of perception, cognition, analysis, reasoning and anticipation skills about the existence and environment of a user, on which, consequently, it can take the appropriate actions. In a smart space, smart things are put into context, form ecosystems that monitor and control our physical environment and our actions.

There are different spaces: smart buildings, such as homes and offices, smart hospitals, hotels and shopping malls, smart cars, and even smart streets. To provide us with the desired comfort, smart spaces want to know everything about us.

Various technologies capture personally identifiable information (PII) and household-level data about citizens - their characteristics, their location and movements, and their activities - link this data to produce new derived data and use it to create profiles of people and places and to make decisions about them.

The smart city infrastructure can read vehicle data using radars, Bluetooth detectors and license plate cameras. Speed, flow and travel times are known in this way and can be associated with the identity of the driver.

9.2.2. Smart infrastructure

Smart cities are based on power and water generation and transmission configurations, transportation frameworks, waste disposal mechanisms, street and home lighting systems, connected healthcare, surveillance, and more.

Large amounts of data are produced by utility companies (electricity, gas, water and lighting use), transportation providers (location / movement, traffic flow), mobile phone operators (location / movement, application use and behavior), travel and smart hotel and accommodation websites (reviews, location / movement and consumption), social media sites (reviews, photos, personal information, location / movement), crowdsourcing and citizen science (maps, local knowledge, urban incidents , climate), government agencies and public administration (services, performance, surveys) and is transmitted through a wireless, mobile and Internet of Things (IoT) infrastructure.

IoT has all the security issues of sensors and actuators, mobile networks and the Internet, namely insecure web interface, insufficient authentication / authorization, insecure network services, lack of transport encryption, insecure cloud interface, insecure mobile interface, in addition privacy concerns (collection of unnecessary data personal information). The data flows generated by the interaction between objects, between objects and individuals, between objects and back-end systems cannot be controlled with classical tools. As the services in a smart city are closely interconnected, if one smart service information system does not provide relevant information to other connected smart services, it can lead to chaotic situations, which can eventually result in total failure**.**

9.2.3. Smart citizens

A smart city is about the relations between the everyday objects surrounding humans and humans themselves, and serving citizens is the main reason of a smart city. In consequence, a smart city will use e-government, will encourage individuals' participation in reporting issues and planning. But, as errare humanum est, people do a lot of mistakes in using the surrounding cyber-physical objects:

- The devices are not configured in an adequate manner, implicit factory settings are used – this is especially dangerous when passwords are involved. Proper authentication settings are not put in place, terms and conditions are not read/understood, there is no knowledge about the data collected by applications and the way of using them;
- Devices are left unattended;
- Stored and transmitted data are treated in the same manner, the sensitive ones are not proper protected;

- People are easily fooled through social engineering, spam emails, data streaming, and other malicious methods.

In addition, as shown above, citizens must self-report various facts about themselves to smart city managers: contact details, financial data, medical data, and emergency data, etc. The collection, processing and transmission of data are not usually explained directly to citizens, and these blindly trusting the way data is used.

Low-quality consent issue appears: in many cases the user does not know the data captured and processed by specific objects. In these situations, it is almost impossible to obtain the consent to data collection / processing required by European law.

9.3. Attacks in Smart Cities

In a smart city, the attack surface is extended, due to the large number of interconnected things, spaces, infrastructures and cyber-physical users. Data security breaches can compromise the entire system, and an infection can easily spread between systems. This, in extremis, can lead to an infection of the city itself, destroying even the physical infrastructure and threatening lives.

Intrusions in SCADA systems can cause disruptions in the exchange of data between control centers and end users. As a result, certain services provided to citizens (access to public health services at critical times, supply of electricity in some areas) will be compromised; certain areas of the city can be blocked by stopping traffic lights, etc. Intruders can also install malware systems in data centers / user devices to obtain confidential information about citizens and use them for criminal purposes.

Some attacks have been recognized since 2014 at the European Union level. Opinion 8/2014 on the Recent Developments on the IoT published by European Union (EU) [29], lists the following situations:

· Inferences derived from data and repurposing of original processing: increasing the amount of data generated by IoT, in combination with modern methods of analysis and cross-matching, allow the use of data for purposes different than the originally established ones. These challenges calls for specific solutions, because, even if the user was comfortable with sharing the original information for one specific purpose, he/she may not want to share this secondary information that could be used for totally different purposes;

- Intrusive identification of behaviour patterns and user profiling: isolated data identified and collected by different objects can be combined to reveal important aspects of the habits, behaviours and preferences of individuals or social groups. Patterns of life and behaviour can be identified in this way. On the other hand, the continuous presence of sensors can put pressure on individuals and can limit their freedom;

- Limitations on the possibility of remaining anonymous when using service: the complete development of a smart city eliminates the citizens' possibility of using services in anonymous mode. The ubiquity of sensors makes it difficult to preserve privacy and poses significant data protection risks;
- Security risks: IoT has numerous security problems, with a risk that every object in the network becomes the target/source of an attack. Risks are therefore more serious than those facing the Internet today. At least two issues should be considered in this case: (1) smart things security, the channels of communication between them and the storage infrastructure and (2) technologies used at different levels of data processing are designed and implemented by different suppliers, without the possibility of standardization and proper protection.[13]

9.4. Security measures in a Smart City

Another Onion Model In order to adequately protect a smart city, a lot of measures provided by various actors are needed. An overall view of these solutions is presented in the figure:

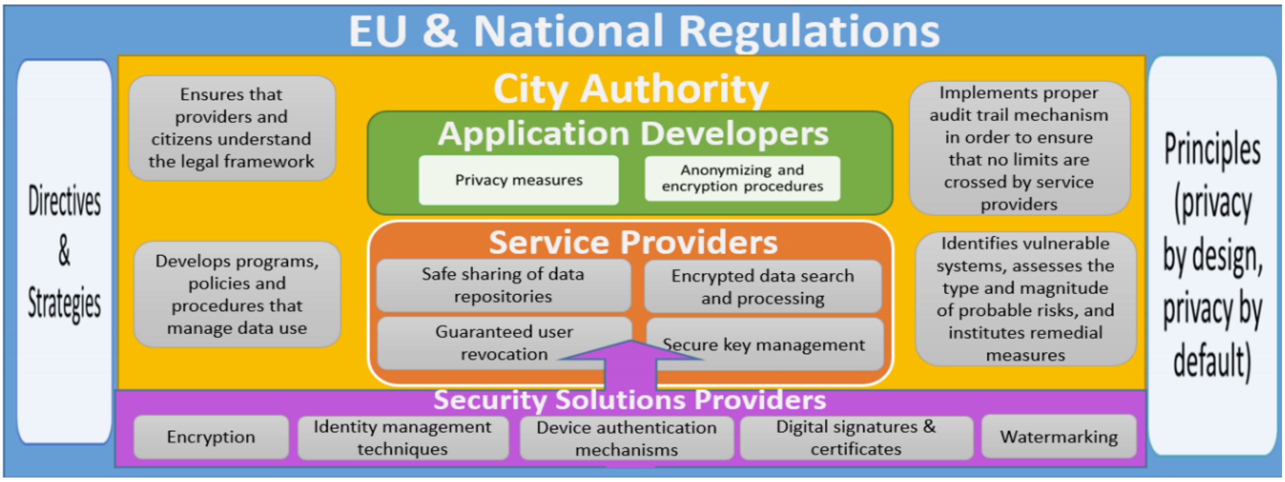


Figure 32 Security measures in a Smart City

The figure shows 5 types of solutions that have to be implemented to have a correct security in the data that flows through the servers of the city.

1. EU & National Regulations: to have a flow and storage of data on different things in the city, it is very important to always do it legally, complying with the regulations required by European organizations.

2. City Authority: it is very important that city authorities respect the privacy of citizens and inform or are authorized by their inhabitants to use certain data. In addition, the authorities are in charge of the security of the data generated by the city.
3. Application Developers: application developers to collect data on city citizens have to put emphasis on privacy measures and anonymize and encrypt both the procedures and the data generated.
4. Service Providers: are in charge of security when sharing encrypted search data, guaranteed user revocation and secure key management.
5. Security Solutions Providers: the latter are in charge of proposing different solutions depending on the problem or depending on the type of security required for each type of data.

9.4.1. European Union regulations

According to the Digital Agenda for Europe, smart environments created by merging the physical and virtual worlds greatly improve the lives of EU citizens. The EU supports the implementation of the smart city concept that allows better public services for citizens, better use of resources and less impact on the environment. In the EU's view, in smart cities, digital technologies translate into better public services for citizens, better use of resources and less impact on the environment.

The EU states that data confidentiality and privacy should play an important role in any smart city development strategy, considering that web-based attacks on IoT increased by 38% in 2015. Otherwise , the introduction of these innovative technologies that access various data about people would not have their consent. Privacy breaches are considered one of the most significant barriers to smart city development by the Alliance for Internet of Things Innovation (AIOTI), an organization founded by the European Commission and several key IoT players in 2015. The "Privacy by design" principle is highly recommended by AIOTI.

According to this concept, privacy protection is built into the earliest stage of technology design. Another important principle is "privacy by default": the controller must implement mechanisms to ensure that, by default, only those personal data that are necessary for each specific purpose of the processing are processed and that, in particular, are not collected. or retain beyond the minimum necessary for those purposes, both in terms of the amount of data and the time of its storage.

The important concepts of trust, security and privacy are treated in strategies as Digital Single Market Strategy, launched in 2015, EU Cybersecurity Strategy, adopted in 2013, European Agenda on Security adopted in 2015. A new term, ePrivacy, was coined for a distinct approach of confidentiality of online PII. In this domain, the main objective is the protection of the confidentiality and the security of individuals' communications in an online environment, which is rooted in the fundamental right to the respect of private and family life. Online privacy is largely approached in the Data Protection Directive and in the ePrivacy Directive (Directive on Privacy and Electronic Communications), which try to ensure safe collection and processing of user data in IoT. Data can be obtained only under strict conditions and for legal purposes, issues that have to be guaranteed by organizations dealing with data collection. The problem is even more important in smart cities, where the volume of data is huge and concerns a wide range of activities in the life of the inhabitants of a city; here, unauthorized access to data can have important negative consequences on large groups of people. To protect them, the European Commission requires telecom operators and Internet Service Providers to report any "personal data breach" to the national authority and to inform the subscriber or individual directly of any risk related to personal data or privacy.[14]

9.4.2. Other stakeholders' actions

A consistent and stable digital architecture must be established. Application developers must specify very clearly the measures they have taken before private and confidential user data is accessed, and the anonymization and encryption procedures used when the data is in transit. Service providers must share their data repositories with other service providers, without compromising their security. Privacy measures, such as searching and processing encrypted data in an untrusted domain, detailed control of shared data, guaranteed user revocation, and secure key management, should be employed to prevent illicit access. to the data.

9.4.3. Security providers

Based on the regulations and stakeholder actions presented above, security vendors must adapt "classic" security methods such as encryption, identity management techniques, device authentication mechanisms, digital certificates, digital signatures, and trademarks. water to the new environment, and making them available to all entities interested in adequate data protection. According to, as attacks continue to increase in sophistication, developing countermeasures remains a challenging and ongoing exercise. Therefore, system-specific attack resistance measures are crucial and devices must dynamically adapt to the situation - scalable security protocols are needed.

9.5. Conclusions

The notions of risk, security and the guarantee of privacy that a smart city should include must be carefully studied. The city authority must be well informed about all problems related to smart things, spaces, services and citizen security; Furthermore, the solution offered by security providers must be known and chosen with the utmost discernment. Each smart city is different, which is why for a correct cybersecurity more in-depth analysis is needed for each vulnerability, attack scenario and adequacy of security measure

10. Risk analysis

A risk analysis offers the possibility of identifying the weaknesses and dangers within a project. Below you will find this analysis, carried out in accordance with the appropriate standards.

Risk analysis - ranking of risks - risk assessment

Project :	**Digital Twins Vilanova**
Date :	**15/04/2021**
Version :	**1**

Estimation of Risk:

A = Non-acceptable risk
B= Acceptable, if accompanied by measures
C= Acceptable

Risk Score (R= P*I)				
0-5	5-10	10-30	30-50	50-100
C	CC	B	A	AA

Kwantificatie van het Risico					
Probability (P)		Impact (I)		Category (C)	
10	Always	10	Sever	S	Schedule
8	Often	8	High	B	Budget
6	Usual	6	Moderate	Q	Quality
4	Possible	4	Low		
2	Almost never	2	Almost No		

Risks for EngineeringTeam

	Description of Risk (Cause - Risk - Impact)	Ranking of Risks C	P	I	P*I	R	Actions to be taken - Decisions for design	Evaluation after reduction P	I	P*I	R
1	We have no specific related background covering "Digital Twins" - Estimation errors - Wrong or not optimal plan of action	Q	4	7	28	B	A. Make a research about Digital Twins concepts B.Wait for more information about the current sensors and data from Neapolis.	1	7	7	CC
1	The proposed program and algorithm does not have the expected impact. - The investment does not come to fruition - Loss of money and time	Q&B	5	8	40	A	A. Find operational applications of the proposed algorithm, before applying it. B. Discuss and consult the found op options with subject specific professors.	1	8	8	CC
1	Covid-19 Pandemic - Less Face to Face contact, less meetings, bad communcition - misunderstandings	S&Q	8	6	48	A	A. a good plan of action B. accurate planning with the required online meetings. + good use of "Trello" - C. use of online sharing platform.	2	6	12	B

Description of Risk (Cause - Risk - Impact)		Ranking of Risks C	P	I	P*I	R	Actions to be taken - Decisions for design	Evaluation after reduction P	I	P*I	R

Risks for Others

Description of Risk		Ranking of Risks C	P	I	P*I	R	Actions to be taken - Decisions for design	Evaluation after reduction Pr	Av	Cl	Se
1	Limited budget - Purchase of quality software, sensors, etc is not possible - Possible problems and difficulties in operationalization.	B & Q	9	5	45	A	A. Research on internet for related software B. Find sponsers or investers for the project.	2	5	10	CC
0					0					0	

Before		
Total Risk	Max Risk (=nb of problem * 100)	Probability and impact Scale (Total Risk / Max Risk) *100
161	400	40%

After		
Total Risk	Max Risk (=nb of problem * 100)	Probability and impact Scale (Total Risk / Max Risk) *100
37	400	9%

	Probability & Impact Scale								
	1	3	4	5	6	7	8	9	10
Rating	LOW	MEDIUM		MEDIUM – HIGH		HIGH		VERY HIGH	
	0 – 5%	5 – 10%		10 – 30%		– 50%		> 50%	
CASE STUDY (Before)						x			

Description of Risk (Cause - Risk - Impact)	Ranking of Risks C	P	I	P*I	R		Actions to be taken - Decisions for design	Evaluation after reduction P	I	P*I	R
CASE STUDY (After)		x									

Table 12 Risk Analysis

The risk analysis above was carried out according to the standard of AP University, which is then further based on the ISO 31000. This standard is an internationally applied standard. The risk analysis will be responsible for the detection, description, severity and solution of the existing risks.

11. Implementation of Digital Twins in Vilanova i la Geltrú

After a great research on how to carry out the process of transforming Vilanova i la Geltrú into a smart city through the Digital Twins concept, we finally know the process that the city must follow for the implementation of this concept. We have divided the process into different parts to make an easier explanation, but the implementation may not follow this order because the creation of the algorithm and the implementation of the Digital Twins concept in the algorithm can be done simultaneously

<u>Phase 1: creation of the algorithm</u>

This phase is the most difficult, important and long because there are many things to do and you cannot have errors since the whole project is based on this phase, the creation of the first part of the algorithm. The way to create the algorithm is described in section 6. Once we have the bases of our algorithm, we need the following to create it:

- Analysed data: one of the needs to manage the algorithm is to feed it with real world data, analysis and predictions of this one. With this information the algorithm will be able to compare the predictions and real world data and detect future problems or inform about the things that could be improved to have a better city .

There is a lot of data that can be analyzed depending on the type of the sensors that the city of Vilanova has. The Digital Twins I project developed a list of sensors that would be necessary for an improvement of the city.

- Solutions: once we have information about the city from the analysed data we need solutions. This is one of the most essential parts. We might think that if we have solutions, why would we need this complex Digital Twins system? The answer is that the algorithm is able to choose the most optimal solution for a problem, reducing the human factor since it can have errors and take more time to propose a solution.
    - On the one hand we need solutions for the anomalies that the algorithm can detect in the future with the live data that the sensors from the city are giving.
    - On the other hand solutions are needed to optimize the resources depending on the live data that is feeding the algorithm.

This part of the phase can be improved by good management of the algorithm's machine learning. Getting him to create solutions for future data analysis.

Our recommendation for this part of the project is commented on in section 7.6 where we indicate that the best option for now to develop the project is the use of Microsoft Azure Digital Twins with the combination of Bentley systems or Smartworldpro from Cityzenith.

Phase 2: implementation of the Digital Twin to the algorithm

Once we have the first part of our algorithm done, we have to implement the Digital Twin to the algorithm. This process is divided into two parts:

1. We must have a storage platform where we have live access to the information provided by all the sensors installed in Vilanova i la Geltrú. Once we have all the real-time information from the sensors, the storage platform or external software has to be in charge of managing the information and classifying the big data in order to get a better reading of the information by the algorithm.

For this process we recommend Iot Hub as a managed service. Azure IoT Hub provides a cloud-hosted solution backend to connect virtually any device and takes care of integrated device management and scaled provisioning of the information it receives.

2. After having the real-time information from the sensors sorted and classified, the next step is to transmit it to the algorithm. The algorithm must be programmed to receive different types of data in real time, to be able to compare them with the historical and analyzed data and finally propose a solution.

Phase 3: launch of Digital Twins

Finally, once we have our algorithm finished with the implementation of Digital Twins on the one hand and the data analyzed with the solutions on the other hand, it is time to start the algorithm.

The software will be analyzing and comparing live information from the sensors every second and when it finds an anomaly or some data that it can optimize, if it can solve it or optimize it, the algorithm will do it and if it is not possible, the operators will be automatically informed.

Example of Azure Digital Twins Integration in Bentley production systems:

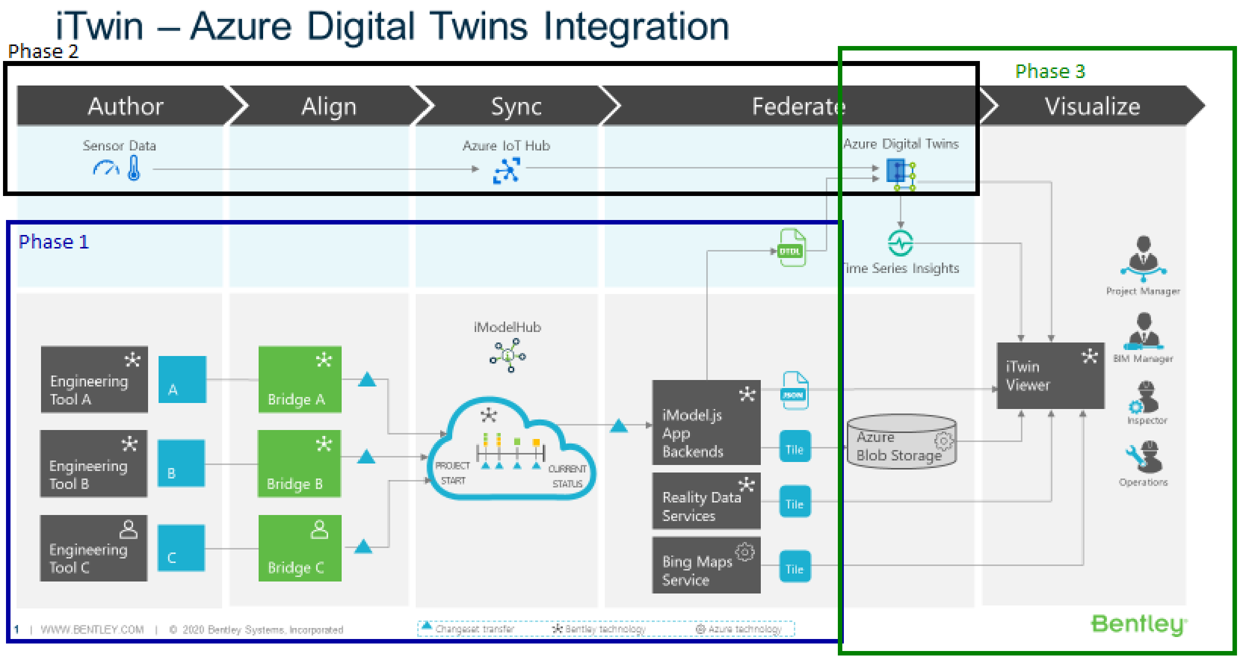


Figure 4 Azure Digital Twin Integration + Bentley systems

We wanted to repeat the figure from section 7.3 for a correct understanding of the process after the explanation of the Digital Twins implementation in this section.The image shows an example of how the Bentley company, using the Microsoft Azure software platform, has implemented Azure Digital Twins in its production systems, achieving excellent results.

In the first phase (surrounded by blue) it is shown how, through a storage platform, they feed the algorithm with data about the engineering processes and possible solutions.

In the second phase (surrounded by black) the real-time information from the sensors is directed to the azure Iot Hub platform for a correct classification of the data and then it is directed to the algorithm produced in Azure Digital Twins.

Finally, in the third phase (surrounded by green), phases one and two come together simultaneously and the algorithm processes the real information and compares it with the database and proposes available solutions and optimizations according to the data it receives.

12. Conclusion and further research

The current research is limited to the investigation of algorithms, software, data management and the problems they entail. Due to limitations such as not too much real-time data, investment costs and time constraints, further elaborations and effective realisations have not been fully achieved. The further research should focus on the processing of the obtained data, their location, type and possible extensions. The proposed software and algorithms should be obtained and worked out. Next, a 3D visualization of the city can be made and the real time data can be read into this. With the help of the chosen algorithms, the necessary analyses can then be carried out and these can be presented graphically in charts.

Further, we want to consider the steps to be taken in the next project Digital Twins III, to complete the transformation of Vilanova and la Geltrú into a smart city:

1. Study the characteristics of this project Digital Twins II.
2. Determine what type of data the project wants to focus on.
3. Study the Big Data from the local sensors provided by the City Council.
4. Propose solutions for possible anomalies in data analysis to feed into the algorithm in the future.
5. Create an algorithm (if necessary, several algorithms) that process, analyze the Big Data and provide a result, thereby proposing a model to manage data for Vilanova i la Geltrú.
6. Apply the proposed model with the Big Data of Vilanova i la Geltrú (if possible, with real ones).
7. Get 3D representations of the proposed Big Data.
8. Load the Big Data and algorithm into Azure Digital Twins software.
9. Obtain predictions and conclusions about the Big Data that feeds the algorithm.
10. Make proposals to improve different topics of the city from the prections.
11. Pros & Cons analysis of the methodology used and of the obtained results.
12. Cost-Benefit analysis of the methodology used and of the obtained results.
13. Propose a methodology to package the proposed methodology, in order it can be exported to other places, to other countries.

In addition, we would like to advise that the team of the next project should have at least 1-2 people who have basic knowledge of machine learning, AI, data mining, neural networks (for example, Convolutional/Recursive/Recurrent neural networks, LSTM etc.) They should be at least beginners in Python and have experience in programming.

Marketing campaign and possible grants

Such a big project like a Digital Twins can need sponsors and grants. This project is made in partnership between EPS (European Project Semester) students and the Neapolis company, which is a company owned by the city council of Vilanova I la Geltrú. So this kind of project needs funding and the best way to receive this funding is to attract sponsors and grants from the government.

Ways to attract founding:

- The project should be clear and easy to understand by everyone and should be very well made to make people understand the importance of the Digital Twins for the community of Vilanova i la Geltrú.
- The project should include all the sensors and the softwares for the implementation
- The project should include all the cost needed for sensors, softwares and other

13. Acknowledgements

We wish to express our sincere appreciation to our supervisors Nora Martinez Antunez and Félix Ruiz Gorrindo, who guided and encouraged us to be professional and do the right thing even when the road got tough. Without their help, the goals and the work at this project would not have been realized.

We wish to thank all the teachers whose assistance was a milestone in the completion of this project. Because they guided us from the beginning until the end of the project, with a lot of advice and support in our journey.

The physical and technical contribution of Escola Politècnica Superior d'Enginyeria de Vilanova i la Geltrú. EPSEVG is truly appreciated by the members of Digital Twins II. Without their support and learnings, this project could not have reached its goals.

14. Bibliography

[1] Thinking about becoming a smart city? 10 benefits of smart cities | Explore Our Thinking | Plante Moran. (2021). Retrieved 6 March 2021, from https://www.plantemoran.com/explore-our-thinking/insight/2018/04/thinking-about-becoming-a-smart-city-10-benefits-of-smart-cities

[2] SmartWorldPro - Cityzenith - The world's most advanced digital twin software solution. (2021). Retrieved 7 June 2021, from https://cityzenith.com/smartworldpro-2/smartworldpro

[3] 3DEXPERIENCity | Marketplace.city. (2021). Retrieved 20 February 2021, from https://marketplace.city/p/3dexperiencity/

[4] Digital Twins: modelado y simulaciones | Microsoft Azure. (2021). Retrieved 20 February 2021, from https://azure.microsoft.com/es-es/services/digital-twins/

[5] Ajunt. Vilanova i la G - Weathercloud. (2021). Retrieved 20 February 2021, from https://app.weathercloud.net/d8705957726

[6] Applications, H., IIoT, E., Environments, C., Code, C., World, M., & Cybersecurity, D. (2021). The right representation of Digital Twins for Data Analytics. Retrieved 7 March 2021, from https://iiot-world.com/industrial-iot/digital-disruption/the-right-representation-of-digital-twins-for-data-analytics/

[7] Digital Twins and Data Analysis. (2021). Retrieved 7 March 2021, from https://www.afcea.org/content/digital-twins-and-data-analysis

[8] Bridgwater, A. (2021). What data does a 'digital twin' run on?. Retrieved 7 March 2021, from https://internetofbusiness.com/data-digital-twin/

[9] Digital Twin – Towards Data Science. (2021). Retrieved 15 April 2021, from https://towardsdatascience.com/tagged/digital-twin

[10] CORDIS | European Commission. (2021). Retrieved 20 February 2021, from https://cordis.europa.eu/article/id/124305-are-smart-cities-ready-for-europes-new-privacy-measures

[11] Appleton, J. (2021). The Importance of Cyber Security & Data Protection for Smart Cities. Retrieved 4 June 2021, from https://hub.beesmart.city/en/strategy/the-importance-of-cyber-security-and-data-protection-for-smart-cities

[12] Smart cities: Is cyber security an issue?. (2021). Retrieved 15 April 2021, from https://www.paconsulting.com/insights/smart-cities-is-cyber-security-an-issue/

[13] Cybersecurity: a smart city imperative. (2021). Retrieved 15 April 2021, from https://www.smartcitiesworld.net/special-reports/special-reports/cybersecurity-a-smart-city-imperative

[14] Smart City Cybersecurity Solutions. (2021). Retrieved 15 April 2021, from https://www2.deloitte.com/us/en/pages/risk/solutions/smart-city-cybersecurity.html

Ltd, R. (2021). Spanish City counts Vehicles to improve Traffic Management System - Retail Sensing. Retrieved 7 June 2021, from https://www.retailsensing.com/people-counting/spanish-city-improves-traffic-management-system-iot-vehicle-counting/

(2021). Retrieved 17 April 2021, from https://www.eea.europa.eu/themes/air/air-quality-index
https://learn.arcgis.com/en/arcgis-book/chapter9/

Global Human Settlement - Download - European Commission. (2021). Retrieved 17 April 2021, from https://ghsl.jrc.ec.europa.eu/download.php

Geltrú, A. (2021). VNG recull dades de circulació de vehicles a la rotonda entre rondes mitjançant sensors. Retrieved 17 April 2021, from https://www.vilanova.cat/noticies/detall?id=99995972

Bagloee S.A., Sarvi M., Patriksson M., et al. Optimization for roads' construction: selection, prioritization, and scheduling // Computer-Aided Civil and Infrastructure Engineering. 2018. Vol. 33, no. 10. P. 833–848. DOI: 10.1111/mice.12370.

Batista G.E.A.P.A., Monard M.C. An analysis of four missing data treatment methods for supervised learning // Appl. Artif. Intell. 2003. Vol. 17, no. 5-6. P. 519–533. DOI: 10.1080/713827181

Kim J., Tae D., Seok J. A survey of missing data imputation using generative adversarial networks // Proc. of the 2020 Int. Conf. on Artificial Intelligence in Information and Communication, ICAIIC 2020. P. 454–456. DOI: 10.1109/ICAIIC48513.2020.9065044.

Mohammadi N., Taylor J.E. Smart city digital twins // Proceedings of the 2017 IEEE Symposium Series on Computational Intelligence (Honolulu, HI, USA, Nov., 27 – Dec., 1, 2017). IEEE, 2017. P. 1–5. DOI: 10.1109/SSCI.2017.8285439.

Ivanov S.A., Nikolskaya K.Yu., Radchenko G.I., Sokolinsky L.B., Zymbler M.L. Digital Twin of a City: Concept Overview. Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering. 2020. Vol. 9, no. 4. P. 5–23. (in Russian) DOI: 10.14529/cmse200401.

Data-driven smart cities: Big Data, analytics and security. 2018. Retrieved 7 May 2021, from https://skelia.com/ articles/data-driven-smart-cities-big-data-analytics-and-security/

DOLL Living Lab | Lighting the future of smart cities. (2021). Retrieved 7 May 2021, from https://doll-livinglab.com

Motion Sensing Outdoor Lighting Control | Comlight. (2021). Retrieved 15 May 2021, from https://www.comlight.no LTD, D. (2018). Why is Open Data Important for Your Business? [online] Medium. Available at: https://medium.com/@DataSciencing/why-is-open-data-important-for-your-business-d77c11ecf948

(2021). Retrieved 15 May 2021, from https://www.slideshare.net/michellewillmers/pros-and-cons-of-open-data-aglobal-south-perspective

Stihler, C. (2021). Open knowledge, AI and algorithms. Retrieved 15 May 2021, from https://blog.okfn.org/2020/03/13/open-knowledge-ai-and-algorithms The problem with Open Data (2012). The problem with Open Data. [online] ComputerWeekly.com.

The problem with Open Data. (2021). Retrieved 1 June 2021, from https://www.computerweekly.com/opinion/The-problem-with-Open-Data Verhulst, S. (2015). The downside of Open Data. [online] The Living Library.

Verhulst, S. (2021). The downside of Open Data. Retrieved 1 June 2021, from https://thelivinglib.org/the-downside-of-open-data/

Kosten für Big-Data-Lösungen: Oracle Appliance versus Eigenbau. (2021). Retrieved 1 June 2021, from https://www.computerweekly.com/de/tipp/Kosten-fuer-Big-Data-Loesungen-Oracle-Appliance-versus-Eigenbau

Measuring Sea Level. (2021). Retrieved 1 June 2021, from https://www.ausmarinescience.com/marine-science-basics/measuring-sea-level/

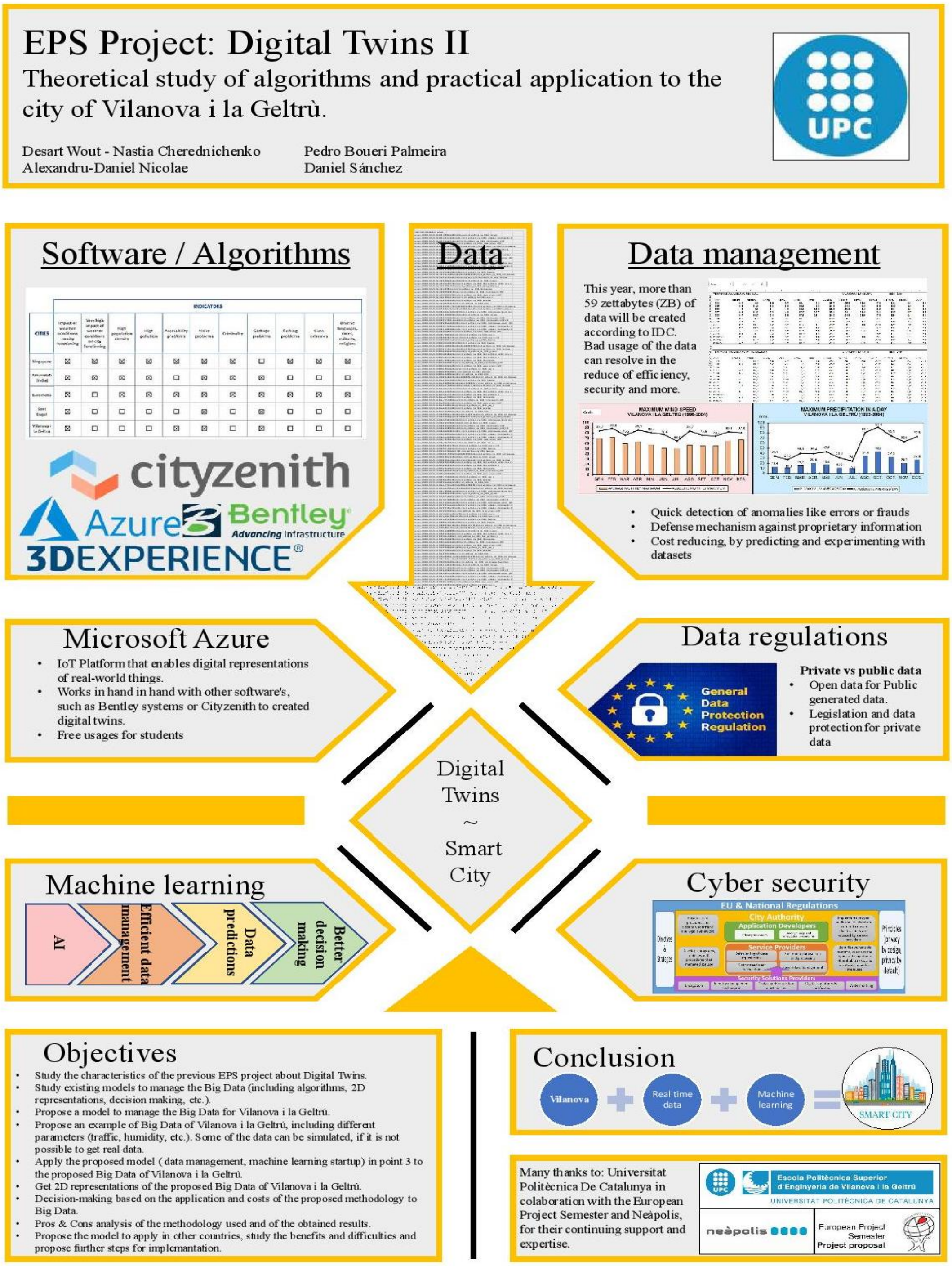
How to install soil moisture sensors—faster, better, and for higher accuracy. (2021). Retrieved 1 June 2021, from https://www.metergroup.com/environment/articles/how-to-install-soil-moisture-sensors-faster/

(2021). Retrieved 1 June 2021, from https://virtualcommissioning.com/top-5-benefits-of-digital-twins/

Anthony Townsend – Blogsmartcities. (2021). Retrieved 1 June 2021, from https://blogsmartcities.wordpress.com/tag/anthony-townsend/

16. Appendix

15.1 Poster



15.2. Project Agreement

PROJECT SCOPE
Digital Twins II

EPS – IDPS 2021

By
Anastasia Cherednichenko
Alexandru-Daniel Nicolae
Pedro Boueri Palmeira
Wout Desart
Daniel Sánchez Molina

19/02/2021

Escola Politècnica Superior d'Enginyeria de Vilanova i la Geltrú
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Escola Politècnica Superior d'Enginyeria de Vilanova i la Geltrú
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Content

Project Scope

Project Name: Digital Twins II. Practical application to the city of Vilanova i la Geltrú

Date: 19.02.2021

1. Project Purpose Statement

The outcome of this project is to develop the practical application of the concept Digital Twins to Vilanova i la Geltrú so that the proposed methodology can be exported to other cities.

2. Background

The company is Neapolis, the public facility in Vilanova i la Geltrú aimed at promoting the knowledge economy, innovation and ICT.

This project is the continuation of the previous EPS project - **Digital Twins I**, in which five international students studied the Digital Twins concept within the summer semester last year and achieved the following goals:

- Study the characteristics of Neàpolis and the city of Vilanova i la Geltrú
- Study the current situation of the Digital Twins concept around the world
- Describe all the advantages and propose the implementation of the Digital Twins concept in Vilanova i la Geltrú
- Propose a plan of phases to implement the concept
- Calculate all the cost of these phases
- Propose an occupation plan, in order the work be made by people and companies of the territory
- Introduce the concept hybridization, in the sense that different professions collaborate generating new transverse professions
- Propose a methodology to package the concept of implementation of Digital Twins in Vilanova i la Geltrú, in order it can be exported to other places/countries
- Investigate the future evolution of the Digital Twins concept around the world

In order to continue working on Digital Twins in Vilanova i la Geltrú and develop the practical application of the concept, we received a full report from the previous project with all the necessary information.

3. Key Stakeholders

Supervisors:

- Félix Ruiz Gorrindo - company supervisor (fruiz@vilanova.cat)
- Nora Martinez Antunez - university supervisor(nora.martinez@upc.edu)

Students:

- Anastasia Cherednichenko - Slovak student of Business Informatics
- Alexandru-Daniel Nicolae - University Politehnica of Bucharest, Romania
- Pedro Boueri Palmeira - Brazilian student
- Wout Desart - Belgian student
- Daniel Sánchez - Spanish student, Universitat Politècnica de Catalunya

4. Objectives / Goals

- Study the characteristics of the previous EPS project about Digital Twins.
- Study existing models to manage the Big Data (including algorithms, 3D representations, decision making, etc.).
- Propose a model to manage the Big Data for Vilanova i la Geltrú.
- Propose an example of Big Data of Vilanova i la Geltrú, including different parameters (traffic, humidity, etc.). Some of the data can be simulated, if it is not possible to get real data.
- Apply the proposed model (data management, machine learning startup) in point 3 to the proposed Big Data of Vilanova i la Geltrú.
- Get 3D representations of the proposed Big Data of Vilanova i la Geltrú.
- Decision-making based on the application and costs of the proposed methodology to Big Data.
- Pros & Cons analysis of the methodology used and of the obtained results.
- Propose the model to apply in other countries, study the benefits and difficulties and propose further steps for implemantation.

5. Other Desirable Requirements

- Make a video of how we managed the project (meetings, visit, teamwork)

6. Related Projects

Since our project has a predecessor, we use the previous final report to continue exploring the Digital Twins field and develop a practical part. This report can be found on our Google Drive, where all files related to the project are stored.

Influence:

- Digital Twins Ⅰ - theoretical part (EPS, februar - june 2020)

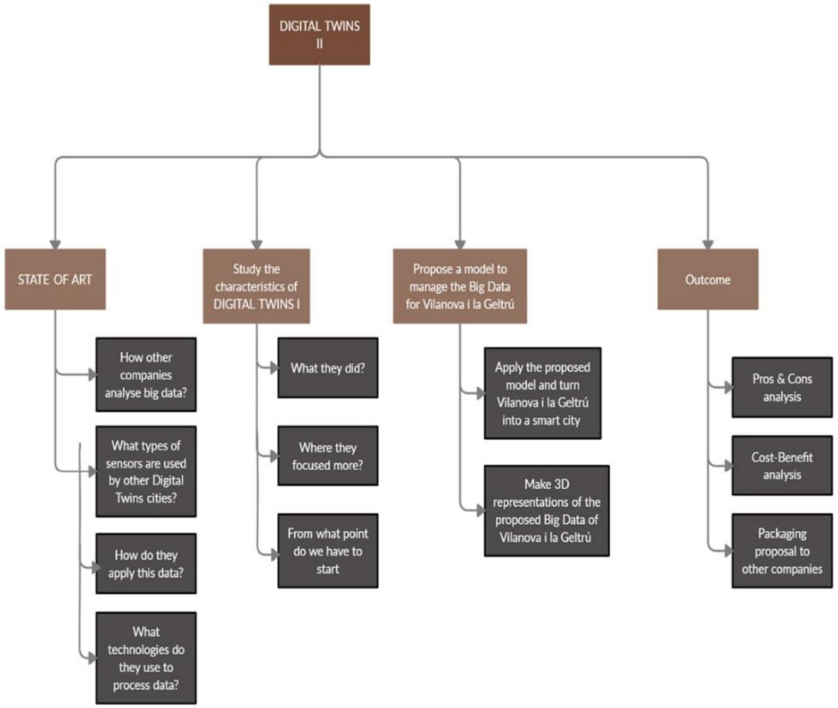
A summary, goals achieved and main points of the previous project will be described in the first chapter of our future report.

7. Approach

To carry out and manage the project, we try to organize mostly face-to-face otherwise online meetings once a week with all team members and with the supervisors once every 1-2 weeks, also we always keep in touch in our WhatsApp group. At the beginning of each week, we discuss tasks for the coming days.

Escola Politècnica Superior d'Enginyeria de Vilanova i la Geltrú
UNIVERSITAT POLITÈCNICA DE CATALUNYA

7.1 WBS



7.2 Planning

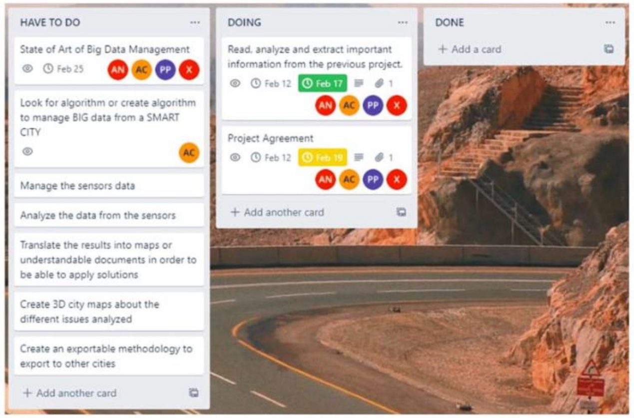
Gantt chart

	Nombre	Duracion	Inicio	Terminado
1	Planning	11 days	8/02/21 8:00	22/02/21 17:00
2	**Study of the characteristics of DIGITAL TWI**	11 days?	**22/02/21 8:00**	**8/03/21 17:00**
3	Current situation in different cities	7 days	22/02/21 8:00	2/03/21 17:00
4	Existing algorithms and smart examples	7 days	22/02/21 8:00	2/03/21 17:00
5	Sensors	7 days	22/02/21 8:00	2/03/21 17:00
6	Processes after data collection	7 days	22/02/21 8:00	2/03/21 17:00
7	Implementation of the Digital Twin concept to Vilar	4 days?	2/03/21 8:00	5/03/21 17:00
8	Conclusion of the Digital Twins I project	2 days?	5/03/21 8:00	8/03/21 17:00
9	**STATE OF ART**	11 days?	**8/03/21 8:00**	**22/03/21 17:00**
10	State of art done in DIGITAL TWINS I	11 days?	8/03/21 8:00	22/03/21 17:00
11	How other companies analyze Big Data?	11 days?	8/03/21 8:00	22/03/21 17:00
12	What types of sensors are used by other Smart C	11 days?	8/03/21 8:00	22/03/21 17:00
13	How do they apply this data?	11 days?	8/03/21 8:00	22/03/21 17:00
14	What technologies do they use to process data?	11 days?	8/03/21 8:00	22/03/21 17:00
15	**Turn Vilanova into a SMART CITY**	36 days?	**22/03/21 8:00**	**10/05/21 17:00**
16	Propose a model to manage the Big Data	16 days?	22/03/21 8:00	12/04/21 17:00
17	Apply the proposed model	26 days?	5/04/21 8:00	10/05/21 17:00
18	Make 3D representations of the proposed Big Dat	11 days?	26/04/21 8:00	10/05/21 17:00
19	**OUTCOME**	21 days?	**10/05/21 8:00**	**7/06/21 17:00**
20	Pros & Cons analysis	11 days?	10/05/21 8:00	24/05/21 17:00
21	Cost-Benefit analysis	11 days?	10/05/21 8:00	24/05/21 17:00
22	Packaging proposal to other companies	11 days?	24/05/21 8:00	7/06/21 17:00

Escola Politècnica Superior d'Enginyeria de Vilanova i la Geltrú
UNIVERSITAT POLITÈCNICA DE CATALUNYA

PM: Kanban Board with Trello

To plan our tasks, we use a Trello board, which contains general information, a list of meetings, as well as lists "HAVE TO DO", "DOING", "DONE". All team members and supervisors have access to this board.



https://trello.com/b/k7JKiBC7/digital-twins-project

7.3 RACI matrix

RACI Matrix

DIGITAL TWINS II

07/03/2021

Deliverable or Task	Status	Anastasia	Alexandru-Daniel	Pedro	Wout	Dani
		ROLES				
		Project Team				
Study the characteristics of DIGITAL TWINS I						
Current situation in different cities		I	R	A	S	I
Existing algorithms and smart examples		S	I	R	A	I
Sensors		R	I	S	I	A
Processes after data collection		I	I	A	R	I
Implementation of the Digital Twin concept to Vila		I	A	R	I	S
Conclusion of the Digital Twins I project		A	S	I	I	R
STATE OF ART						
State of art done in DIGITAL TWINS I		I	R	I	S	A
How other companies analyze Big Data?		S	I	R	A	I
What types of sensors are used by other Smart Ci		R	I	S	I	A
How do they apply this data?		I	I	A	R	I
What technologies do they use to process data?		I	A	R	I	S
Turn Vilanova into a SMART CITY						
Propose a model to manage the Big Data		R/A	R/A	R/A	R/A	R/A
Apply the proposed model		R/A	R/A	R/A	R/A	R/A
Make 3D representations of the proposed Big Data		R/A	R/A	R/A	R/A	R/A
OUTCOME						
Pros & Cons analysis		R/A	R/A	R/A	R/A	R/A
Cost-Benefit analysis		R/A	R/A	R/A	R/A	R/A
Packaging proposal to other companies		R/A	R/A	R/A	R/A	R/A

D	Driver	Assists those who are responsible for a task.
R	Responsible	Assigned to complete the task or deliverable.
A	Accountable	Has final decision-making authority and accountability for completion.
S	Support	Provides support during implementation.
C	Consulted	An adviser, stakeholder, or subject matter expert who is consulted before a decision or action.
I	Informed	Must be informed after a decision or action.

Escola Politècnica Superior d'Enginyeria de Vilanova i la Geltrú
UNIVERSITAT POLITÈCNICA DE CATALUNYA

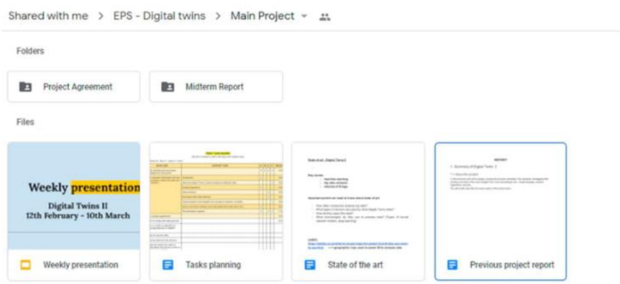
7.4 Team rules

In order to make our team work and avoid misunderstandings, we adhere to the following rules:

1. Equal work
2. Be in touch
3. Respect for all team members
4. Responsibility for your work
5. Ask if you don't know
6. Always reply to messages in WhatsApp
7. Say if you don't like something
8. Offer help if you can
9. Friendliness
10. Positive attitude

7.5 Data Storage Management

To organize our Data Storage, we use a shared Google Drive, link to which you can find below. We have a folder "Main Project", where all files related to our project are stored. We also have a separate folder for each main task to make navigation easier.



https://drive.google.com/drive/u/1/folders/17d5KX9SK7baotmsvzJBhIaz-ED-R_bG2

8. Timeframe & Milestones

Project milestones allow us to track the progress of the projects from start to finish. In this way, we will know if we get behind. Far in advance before the final deadline and in this way we will be able to adjust our plans or expectations to stay on target.

Project timeframe will help to set clear directions and priorities. These timelines shape the entire project and keep everyone informed and aligned at every stage of the project.

Timeframe & Milestones	
Study the characteristics of DIGITAL TWINS I	22.02.2021-08.03.2021
State of art	01.03.2021-22.03.2021
Intensive seminars	08.02.2021-05.03.2020
Submission Midterm Report	15.03.2020
Prepare for Midterm Defenses presentation	15.03.2021-21.03.2021
Midterm Defenses	22/23.03.2021
Easter Holidays	29.03.2021-05.04.2021
Prepation of the Final Report	05.04.2021-25.05.2021
Group review of the Final Report before sending it to supervisor	25.05.2021-01.06.2021
Submission Final Report for final review to supervisor	01.06.2021
Submission Final Report, Article, Poster, Video	08.06.2021
Final defenses	14/15.03.2021

9. Inclusions & Deliverables

Inclusions:

We have to convert Vilanova i la Geltrú into a Smart City, based on the concept of Digital Twins, making the life of citizens much easier and more comfortable. By placing sensors all over the city of Vilanova i la Geltrú, who can collect data from different kinds of environments from the city.

Deliverables:

At the end of the project we have to deliver a way to manage Big Data, collected by the sensors from the city of Vilanova i la Geltrú. Also to make a 3D representation of the proposed Big Data of Vilanova i la Geltrú.

10. Exclusions

Items that are not within the project boundaries:

- testing the model on dates from another city
- development of several different models
- installation of sensors
- connecting the developed model to the city sensors

11. Critical Success Factors (CSF)

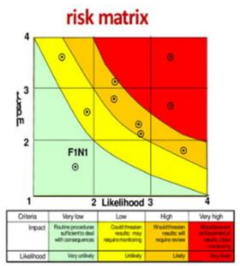
- students are in touch with the supervisors each week and set the correct weekly tasks
- students have sufficient knowledge or the opportunity to acquire it in such fields as Big Data, deep learning, neural networks and Python programming
- students are correctly informed about what types of sensors they have to work with, with what datas, with what required data standardization and with what expected data output to create a suitable model
- students are given provided data from the sensors of the city of Vilanova
- students have free access to programs/platforms for 3D modeling

12. Assumptions

- We assume that we need to create an algorithm for managing data for each data type
- We expect to get free access to programs and modules for the development of the Digital Twins concept

13. Constraints

Efficient project management is one of the most important tools in organizations, as people are searching for more ways to improve their operations. The most common constraints in our project are going to be **time** and **cost**. Big data is an enormous idea that comes with a lot of difficulties. The search for the right equipment, algorithms and implementation. All these factors take up alot of time and money.



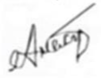
Furthermore our project has a certain risk factor, by comparing the probability and the impact. We can say that if the project doesn't have the impact we hoped for, the investment won't pay off. Therefore the increase in the risk.

Further, the project is getting realistic during a pandemic by international students.

Most conversations will take place online, this because the current situation can't allow non-presential classes. **The lack of physical contact** will not positively affect the appointments or the overall progress of the project.

14. Sign Offs

This is an agreement between the project sponsor and project manager to the high level requirements of the project as known to this point.

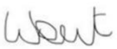


Anastasia Cherednichenko

Daniel Sanchez Molina

Alexandru-Daniel Nicolae

Pedro Boueri Palmeira



Wout Desart

16.1. Digital Twins glossary

This glossary is intended for the next students who will develop the digital project twin 3 in order to inform previously about terms and concepts of extreme importance when developing a digital project twin

The information present at this point was obtained from one of the companies currently working with digital twin projects, in this case, the company is called "digital twin consortium" and can be accessed through the website for more detailed information (https://www.digitaltwinconsortium.org)

Aggregation

In software architecture, aggregation implies gathering, copying, and possibly transforming information from multiple systems into a single centralized system.

Computational Representation

A computational representation is an **executable** digital representation consisting of computational algorithms and supporting data representing some subject matter from a dynamic perspective.

include simulations and predictive analytics. Examples of computational representations include Finite Element Analysis (FEA) models, Machine Learning models, and various other kinds of simulations based on mathematical equations, including those describing laws of physics and engineering.

Cyber-Physical System

A system consisting of physical and digital systems integrated via networking.

A digital twin considered together with its physical twin is an example of a cyber-physical system.

A physical device with a closely integrated digital control system is another example of a cyber-physical system. Such a cyber-physical system may be considered a real-world entity represented by a digital twin.

Data Interoperability

Data Interoperability is a functional digital twin subsystem that supports its integration representation/function by aggregating or federating data from digital systems that may be considered "inside" or "external to" the digital twin system.

Data Model

A data model is a model of data that describes its structure, datatypes, and meaning.

Alternate terms

- "Ontology" is sometimes used as a synonym for "data model", particularly for data-modeling ontologies
- "Schema" is sometimes used as a synonym for "data model". DDL defines database schemas. OData uses CSDL (Common Schema Definition Language). RDFS (Resource Description Framework Schema) is sometimes used in conjunction with OWL to define data models using a triple-graph approach.
- "Information Model" is sometimes used as a synonym for "data model", though some parties make subtle distinctions between "information model" and "data model", such as taking it to imply a data model that lies on the "conceptual" end of the implementation spectrum.

The various synonyms of "data model" and the fact that some ontologies can be used as data models can be a significant source of confusion in the design and development of stored representations for digital twin systems.

Data Modeling in Digital Twin Systems

Data models and ontologies play various functional roles in digital twin systems.

A persistence data model uses a data modeling language to describe data structures and types in a way that is compatible with the persistence technology. If the persistence technology is a relational database, this will be a SQL DDL "schema". Other persistence technologies will use different data modeling paradigms and data modeling languages.

Service interfaces embody a logical data model that describes the data structures and types used by an API or protocol. This data model may differ from the lower-level persistence data model, with the API implementation handling the mapping between the two. Examples include OData's CSDL (for entity-relationship style REST APIs) and GraphQL Schema (for property-graph style APIs). Other common ways of defining the "logical" data model of an API or protocol include JSON Schema and XML schema. Non-web-based APIs model data in programming-language-specific ways.

Conceptual data models are intended to be relatively free of the implementation-specific functional, performance, and scalability concerns of service interfaces and stored representations so that they can focus on describing the real-world concepts in the domain of the digital twin. These data models may or may not be machine-readable, as they may only be used to guide the development of the GUI and lower-level data models.

Conceptual data models may be used in a digital twin system that is integrating information from multiple stored representations, each with its own persistence and logical data models. The digital twin system may have an integration representation/function that provides access to data using a conceptual data model that is compatible with the overall ontology of the digital twin system. . A variety of data modeling languages can be used for this function, including those for data-modeling ontologies.

Because the integration representation/function may employ both aggregation and federation integration strategies, data modeling languages compatible with integration-friendly service interfaces like OData, SPARQL, and GraphQL may be useful. These support both access to aggregated integration representations and mapping to physically-separate federated stored representations.

Most data modeling languages can be used in many of these functional roles, and roles can be combined. A single data model could be used at the conceptual, service interface, and persistence levels if system requirements can be met. The data-modeling ontology of the integration service interface may be used as the overall ontology of the digital twin system and the data model of the integration representation/function.

Data Modeling Language

A data modeling language is a lexical or graphical language used define the data structures and data types of a data model.

Examples of graphical data modeling languages include:

- UML (Unified Modeling Language): a general modeling language that can be used for data modeling.
- ERD (Entity Relationship Diagrams): a data modeling language well-suited for relational databases.

Lexical data modeling languages include:

- SQL Data Definition Language (DDL): The standard language for relational database schemas.

- Digital Twin Definition Language (DTDL): A property-graph data-modeling language with support for integration with IoT via "telemetry" properties. DTDL was created by Microsoft and open-sourced.
- GraphQL Schema: A property-graph data-modeling language used with GraphQL to support federation of data from multiple sources and purpose-specific views of data. GraphQL was created by Facebook and open-sourced.
- Web Ontology Language (OWL): A triple-graph data-modeling language that is part of the Semantic Web Stack.
- EXPRESS: A entity-relationship data-modeling language used by STEP for modeling geometry and IFC for modeling built assets.
- Many more examples can be found here.

Data-modeling Ontology

A data-modeling ontology is an ontology that is also a conceptual data model.

Data-modeling ontologies describe real-world universals using data structures and datatypes in a particular data-modeling paradigm. Ideally, they also include precise human-readable definitions of the universals that they are describing.

An incomplete list of examples of data-modeling ontologies relevant to digital twins includes:

- DTDL
- OWL

Data Modeling Paradigm

A data modeling paradigm is an approach to data modeling embodied in a set of fundamental constructs that govern how data is structured and typed.

Digital Representation

A digital representation is a representation (aka "model") of some subject matter consisting of structured digital information and/or computational algorithms.

Before computers, we represented/modeled physical things using clay, wood, or plastic models or drawings on paper. Nowadays, we use digital representations, which may include 3D models, drawings, or lexical databases of various sorts.

Digital System-of-Systems

A digital system-of-systems is a digital system comprised of one or more other digital systems

Digital Thread

Digital thread is a mechanism for correlating information across multiple dimensions of the virtual representation, where the dimensions include (but are not limited to) time or lifecycle stage (including design intent), kind-of-model, and configuration history.

Digital Twin

A digital twin is a virtual representation of real-world entities and processes, synchronized at a specified frequency and fidelity.

Digital Twin Systems transform business by accelerating holistic understanding, optimal decision-making, and effective action.

Digital Twins use real-time and historical data to represent the past and present and simulate predicted futures.

Digital Twins are motivated by outcomes, tailored to use cases, powered by integration, built on data, guided by domain knowledge, and implemented in IT/OT systems.

Digital Twin Applications

Digital twin applications are software applications that leverage digital twin services and/or service interfaces.

Digital Twin Platform

A digital twin platform is a set of integrated services, applications, and other digital twin subsystems that are designed to be used to implement [digital twin systems].

Digital Twin Services

Digital twin services are functional subsystems of a digital twin system that provide value by leveraging the digital twin.

Examples of digital twin services include:

- Visualization services for web, mobile, desktop, Virtual Reality (VR), and Augmented Reality (AR) devices.
- Analysis services of various kinds.
- Many other kinds of services that will be enumerated at a later date.

Most digital twin services provide a service interface for use by other services and digital twin applications.

Digital Twin System

A digital twin system is a system-of-systems that implements a digital twin.

Digital Twin System Feature

A digital twin system feature is a functionality of a digital twin system.

Digital twin system features are generally implemented by subsystems of the digital twin system

Digital Twin Use Case

A digital twin use case is a use case in which digital twins may be used to accomplish desired outcomes.

Digital Twin use cases are designed to improve outcomes, provide requirements for digital twin systems.

Federation

Federation is an integration strategy that involves coordinated access to data repositories without making centralized copies of them.

Federation only implies gathering enough centralized "index" information to use the federated systems in a coordinated manner, without copying the bulk of their data. If transformation of data is required, it is performed on-the-fly.

Integration Representation/Function

The integration representation/function is a subsystem of a digital twin system that provides integrated and semantically-aligned access to the stored representations of the digital twin system.

The integration representation/function may comprise a stored representation for data aggregation and functions for data federation.

The integration representation/function is closely associated with digital thread.

Integration Service Interface

An integration service interface is a service interface for the integration representation/function of a digital twin system.

Integration Strategies

Integration Strategies (in digital twin systems) are strategies for integrating data/information to achieve a cohesive virtual representation.

In software architecture, there are two main patterns for data integration: aggregation and federation.

Interventional Frequency

An interventional frequency is a synchronization frequency characterizing how often interventions in reality occur in order to synchronize reality with the state of a stored representation.

IT/OT Platform

The IT/OT Platform of a digital twin system is the set of Information Technology and Operational Technology infrastructure and services on which the subsystems of a digital twin system are implemented.

Subsystems of the IT/OT Platform include:

- A software platform and tooling stack.
- Platform APIs
- Orchestration of low-level infrastructure
- Compute, storage, and networking infrastructure

These systems could be cloud-based, on premises, embedded, mobile, distributed, etc.

Management and Automation

Management and Automation is a functional digital twin subsystem that supports general management and operation of the digital twin system.

Model

A representation of some subject matter modeled in some medium from some modeling perspective.

The subject matter could be a car, a city, a supply chain, or a factory and its surroundings.

The modeling medium could be clay, wood, plastic, mathematical equations in a spreadsheet, or ink-on-paper. Digital twins are mostly concerned with the "digital" medium--information stored as bits and bytes that can be loaded into a computer's working memory.

Observational Frequency

An observational frequency is a synchronization frequency characterizing how often observations of reality are made and reflected in a stored representation.

Ontology

An ontology is a representational artefact that describes universals and certain relations among them in a domain of interest.

"Universals" are types or classes of entities, in contrast to "particulars" which are instances of entities.

An ontology can be written in any number of natural or formal languages and a given ontological understanding can be represented in more than one language.

A human-readable ontology is used to assist people's understanding of and communication regarding the ontology's domain.

A machine-readable ontology can be read by software to associate the semantics defined in the ontology with data. This may make the data more meaningful to humans. It may also facilitate data interoperability if data structures from different systems are mapped to terms from a common ontology.

Machine-readable ontologies should also include detailed human-readable definitions of their terms in order to be used effectively for meaningful semantic mapping.

Ontologies generally do not specify data structures or data types used to represent particular entities, but data-modeling ontologies do.

An ontology can be used in a digital twin system to provide consistent semantics and support data integration among different [stored representations

Also see wikipedia's entry for ontology (information science) and ontology (philosophical).

Physical Twin

A physical twin is a set of real-world entities and processes that corresponds to a digital twin

The physical twin is the subject matter modeled by a digital twin.

The physical twin may constitute physical systems of interest and their environment, interactions, and processes.

Real-to-virtual synchronization

Real-to-virtual synchronization (aka observational synchronization) is synchronization that causes a stored representation to reflect new observations of the real world.

In other words, it is the process of mirroring the real world in the virtual representation, based on observation of the real-world.

Real-World

In the context of digital twins, "real-world" refers to the subject matter of a stored representation or digital twin (which can be considered as a kind of stored representation).

Security, Trust, and Governance

Security, Trust, and Governance are a set of foundational concerns for a the functional subsystems of a digital twin system.

The concerns include privacy, security, safety, resilience, and reliability.

Service Interface

A system's service interface is a digitally addressable endpoint that implements a protocol through which other systems and services may interact with the system.

Simulation Modeling Language

A simulation modeling language is a lexical or graphic language used to define simulation models for computational representations.

Examples include Modelica, Simulink, etc. Simulations can also be coded in programming languages such as FORTRAN, C, Lisp, etc. Not all computational representations are defined via simulation modeling languages, because simulation techniques and technologies vary widely.

Subsystems of a Digital Twin System

A subsystem of a digital twin system is a system that implements a digital twin system feature

Stored Representation

A stored representation is a digital representation consisting of stored structured information, representing states of some subject matter.

A stored representation can be queried, in contrast to a computational representation, which must be executed to produce output.

Synchronization

Synchronization is the process of causing the virtual representation to more-closely match the real-world or cause the real-world to more-closely match the virtual representation of a desired state.

Synchronization is implemented via some synchronization mechanism.

Synchronization Frequency

Synchronization Frequency is a frequency characterizing how often synchronization occurs

The frequency will not be uniform for a digital twin. It may vary per stored representation or even within the stored representation.

Synchronization Mechanism

A synchronization mechanism is a mechanism through which synchronization is implemented.

Use Case

A use case is a set of circumstances or a scenario for the use of something.

Virtual

Not physically existing as such but made by software to appear to do so.

Virtual Representation

A virtual representation is a complex, cohesive digital representation composed of stored representations, computational representations, and supporting data which collectively provide an information-rich "virtual" experience of their subject matter.

The integration representation/function of a digital twin system "virtually" joins information of various kinds together into the cohesive, multi-faceted, representation of reality that we call a "virtual representation".

Virtual-to-real synchronization

Virtual-to-real synchronization synchronization that intervenes in the real world to make it more-closely match a stored representation of a desired state.

In other words, it is the process of mirroring the virtual representation into the real world, through some intervention in the real world.

16.2. Contact Reports

16.2.1. Report1

CONTACT REPORT

Av. de Víctor Balaguer, 1, Vilanova i la Geltrú, Barcelona| 08800 • 938 96 77 01 Tel. | www. upc.edu.com

Meeting Date: 17/03/2021

Report Date: 16/04/2021

Company/Uni Attendees:
Amedee Beylemans

Communication Type: Email

Location: Online

Contact Report Number: CR1

Engineering Team Attendees:
Wout Desart

Third Party Attendees: /

Subject: IMEC MICROSOFT AZURE

PURPOSE:

Aids in researching and contacting for the Microsoft Azure software program. With the company in question IMEC.

SUMMARY:

The necessary contacts with IMEC cannot be obtained. IMEC is cooperating with another university (KUL). The city of Antwerp has started another IOT project, which goes under the name "City of Things". For this I have obtained two new contacts.

The further help for obtaining Azure, did not work out. The software is not part of the package that the university can offer.

GATHERED INFORMATION:

The project between the city and the faculty of applied engineering electrical engineering and electronics continues under the name "City of Things.

Obtained contacts:

- Mr. Patrick Cools - senior lecturer

- Mr. Prof. Dr. Walter Daems

- dhr. Ive Beeckmans to help obtain software

DECISIONS/FURTHER ACTIONS:

Make contact with the addresses obtained in order to expand the research and obtain the necessary information and software.

Figure 33 Contact Report 1

16.2.2. Report2

CONTACT REPORT

Av. de Víctor Balaguer, 1, Vilanova i la Geltrú, Barcelona| 08800 • 938 96 77 01 Tel. | www. upc.edu.com

Meeting Date: 17/03 – 28/04

Communication Type: Mail

Report Date: 28/04/2021

Location: Online

Client Attendees: Amedee Beylemans

Contact Report Number: 2

Agency Attendees: Wout Desart

CC: X

Third Party Attendees: /

Subject: Access Microsoft Azure + Information IoT project Antwerp

PURPOSE:

The purpose of the communication is to gather more information about the use and choise of Microsoft Azure, to get more contacts and to try to get an free access to the software platform.

SUMMARY:

- **Me**: asked for contact information with IMEC (Belgian company using Microsoft Azure) & asked to get a free license for Microsoft Azure.
    - **Amedee**: IMEC has partnership with a university in an other city (Gent), doesn't have direct contacts with company. Gathered information about an other IoT project in Antwerp. Made two new contacts for these projects.
    - **Amedee**: I got an contact (Ive Beeckmans) for the IT department, to get a license key for Microsoft Azure.
- **Me**: Send mail to Patrick Cools and Walter Deams, asking them for more information about (the choise) of Microsoft Azure and if they have any doubts.
    - **Patrik**: Not his area of expertise, got redirected to Walter
    - **Walter**: Got redirected to Peter Hellinckx, direct partner with "City of things" IoT project.
- **Me**: Send mail to Ive, regarding access to Microsoft Azure.
    - **Ive**: Not part of ICT department anymore. Got redirected to Korneel Stassen.
    - **Ive**: There is an trail version. Told us to use this.

- **Me**: Send mail to Korneel Stassen, regarding access to Micorsot Azure.
  - **Korneel**: No Rely Yet
- **Me**: Send mail to Peter Hellinckx, regarding Microsoft Azure.
  - **Peter**: No Repy Yet

GATHERED INFORMATION:

Contacts:

- Patrik Cools → Head Professor ElektroMechanica University Antwerp
- Walter Deams → Head of industrial engineering department (Vicedecaan)
- Peter Hellinckx → Part of Iot project “City Of Things” Antwerp
- Ive Beeckmans → Ex-part of the ICT department of AP
- Korneel Stassen → Part of ICT department AP

DECISIONS/FURTHER ACTIONS:

- Waiting on further Information

Figure 34 Contact Report 2