IMPLEMENTATION OF DIGITAL TWINS IN VILANOVA I LA GELTRÚ

Wout Desart, Nastia Cherednichenko, Pedro Boueri, Alexandru-Daniel Nicolae, Daniel Sánchez

***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 are that it significantly increases the city 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 continue researching about Digital City Twins and explore the Big Data from the city sensors of Vilanova i la Geltrú.**

***Keywords*—** Big Data, Algorithm, Data Analaysis, Digital Twin

I. Introduction

Nowadays, new technologies are part of our daily lives making life simpler and easier. This is the reason why Neapolis co with 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, which is a replica of the physical world. The development of this concept is through a program based on Internet of things (IoT) powered by sensors, which are capable of, among other things, detecting problems, preventing problems and costs and trail and shape predicted outcomes in 3D representations of the city

Continuing with the project carried out by the European Project Semester (EPS) students last semester, DIGITAL TWINS I, where they conducted 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 what they did and implement the digital twins concept in the coastal city by proposing and applying a model to manage Big Data. [1]

II. Project goals and objectives

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

Some of 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

III. About the previous project Digital Twins I

The previous project Digital Twins I describes the main concepts associated with Digital Twins, history and current situation; advantages/disadvantages of implementing Digital Twins to smart cities; general about 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.

IV. 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 and 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. All the factors will be enormously dependent on underlying factors and will therefore fluctuate. As 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 wellbeing 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.

V. ALGORITHM

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.

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 behavior of the physical assets, and (4) simulation of the smart city process and its optimization according to a set of optimization constraints.

Once we have the basis of how our algorithm will be like, 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.[2] [3]

VI. PLATFORMS/SOFTWARE FOR DEVELOPING DIGITAL TWINS

After a great research and comparison between different software and products, our conclusion is the use of Microsoft Azure, the cloud-based computing platform.
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)

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.

VII. 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. Sensors listed 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

VIII. DATA CLEANSING & FOGGY COMPUTING

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 40 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.

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.

IX. DATA

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.[4]

i. Example of analyzed Big Data

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.

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 1 Example of the distribution of temperature data in VNG in Excel [5]

From the data previously shown in Excel we can extract graphs such as the following:

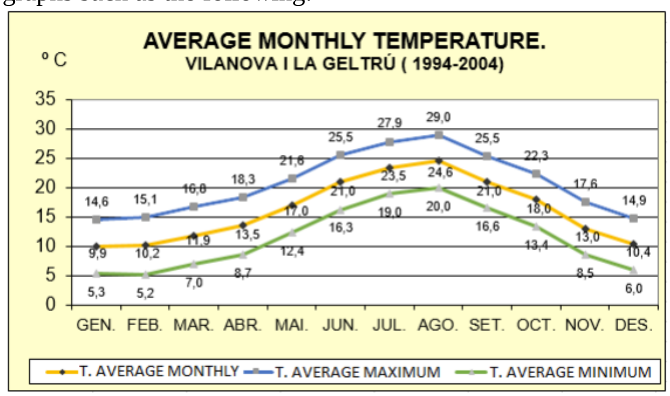

Figure 2 Average Monthly Temperature in VNG (1994-2004)

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.

ii. Example of bad Big Data

Previously we have seen an example of Big Data in an excel file generated by different sensors. In this example 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	condensado	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	condensado	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 3 Example of bad Big Data in EXCEL

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.

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.

X.

XI. Data security

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 [6][7]

XII. Implementation of digital twins in Vilanova i la Geltrú

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.[8]

Phase 1: creation of the algorithm

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
- 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.

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.
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.[9]

XIII. 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) and get 3d representations of the proposed Big Data.
7. Load the Big Data and algorithm into Azure Digital Twins software and Obtain predictions and conclusions about the Big Data that feeds the algorithm.
8. Make proposals to improve different topics of the city from the prections.
9. Pros & Cons analysis of the methodology used and of the obtained results and Cost-Benefit analysis of the methodology used and of the obtained results.
10. Propose a methodology to package the proposed methodology, in order it can be exported to other places, to other countries.

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.

REFERENCES

[1] 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.

[2] 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. H. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1985, ch. 4.

[3] 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.

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

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

[6] Appleton, J. (2021). The Importance of Cyber Security & Data Protection for Smart Cities. Retrieved 7 June 2021, from https://hub.beesmart.city/en/strategy/the-importance-of-cyber-security-and-data-protection-for-smart-cities

[7] 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

[8] 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

[9] 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/