Open Data for simulation to determine the efficient management of parking spaces in Smart City

Iwona Chomiak-Orsa, Piotr Domagała, Andrzej Greńczuk,
Wojciech Grzelak, Artur Kotwica, Kazimierz
Perechuda Wroclaw Uniwersyty of Economics and Business, Komandorska 118/120, 53-345
Wrocław Poland;
E-mail:{iwona.chomiak-orsa, piort.domagała, andrzej.greńczuk,
wojciech.grzelak, artur.kotwica, kazimierz.perechuda}@ue.wroc.pl

Abstract—**The problem of optimization and effective management of parking spaces is one of the main problems faced by modern cities. Therefore, in creating Smart Cities solutions, more and more attention is focused on the possibilities of using advanced ICT tools to improve these processes. According to the authors of the article, such a method can be the creation of the so-called Digital Twin. In order to present the simulation possibilities that can be achieved by using a Digital Twin, the authors identified the chances of obtaining the data resources necessary for creating the simulation. Identification and evaluation of data sources had a character of a pilot study and referred to four car parks located in Wroclaw. On the example of one of them, an attempt was made to create an indicator of the type of traffic. The theoretical considerations and research presented in the paper are elements of research on creating Smart City solutions carried out by the team of authors.**

I. INTRODUCTION

NOWADAYS, we can observe the growing use of ICT (Information-Communication Technology) in various areas. ICT plays an important role in our lives and the proper functioning of organizations. It is used not only to facilitate communication, data storage, but more and more in every possible way. The current development of technology has focused on the use of mathematics, statistics and econometric models in analytics and decision-making processes. The development of tools and new technologies has resulted in the development of new tools that help reflect reality and make changes on it before they are implemented in "real life".

The created tools allow not only to model a potential scenario (through ongoing verification of "combined" elements), but also to recreate it in a virtual environment, change and assess the quality of these changes. Most people make decisions and/or create new solutions based on the simulation of a specific thing/phenomenon. The word "simulation" itself comes from Latin and means "pretending". It can also be said that simulations play a bigger role today than a few years ago. With the help of simple simulations, we can quickly and easily generate the output data, based on the input sample. Such results can then be further analyzed (including the use of data drilling algorithms) to detect appropriate relationships between the data and create specific models.

Simulations are used in various fields such as economics, including business, mathematics, computer simulation games, engineering sciences. Programs supporting the creation of simulations can be domain-specific as well as general use. This means that not every program will be able to implement a specific project.

The authors in this article continue their research in the field of efficient parking space management in Smart Cities. During the analysis of the literature, it was noticed that until now computer simulation was not used. A "simulation" was used, the aim of which was to create a prototype of the system. These prototypes either generated data based on a specific algorithm or tried to create a basis for a simulation model.

The article consists of the following parts. The first part discusses the role of simulation, Digital Twin and open data in improving Smart City processes. The second part deals with the problems of modern parking spaces in cities. The third part shows how open data can be used in the simulation of parking spaces in cities on the example of four parking lots in Wrocław. The whole thing ends with a discussion and conclusions in the area of open data usability in the simulation process.

II. SIMULATION BASED APPROACH TO STREAMLINED SMART CITY PROCESSES

Simulation is an approximate reproduction of the phenomena or behavior of an object using its model. A special type of model is a mathematical model, often written in the form of a computer program. The concept of simulation was borrowed from traditional language learning. To "simulate" means to look similar to other people or to copy the behavior of such people. In the context of computer simulation, we are talking about copying the operation of the entire system or copying specific situations with the use of a computer program.

More precisely, "computer simulation" can be called a numerical method used to carry out experiments on specific types of mathematical models that characterize, using a

digital machine, the operation of a complex system over an extended period of time.

From these definitions it can be concluded that:

- computer simulation is a method based on conducting research on dynamic models that discuss the existing or developed systems,
- the research reason for the computer simulation is to obtain information about the work of the analyzed system over time,
- computer simulation uses a computer program as a work tool while carrying out the research objective, which is the official representation of the model of the analyzed system.
- computer simulation as a method is a system of consciously selected research activities, ie the structure of phased works leads to the achievement of the set research goal.

Conducting a simulation enables the analysis of the process in various variants, which are verified in a virtual way, thus not affecting the activity of the process in real time. However, based on well-developed control parameters, consistent with the actual state, it can be said with high probability that the analyzed process variant has a chance to be implemented in the economic reality [7]. Each simulation requires the definition of basic principles [4]:

- in the case of complex processes subject to simulation, it is necessary to properly select the tool used for the simulation and detailed modeling of the parameters of the analyzed process and the system in which it operates, defining the input data and defining the goal,
- in the case of flexible processes subject to simulation, it is necessary to frequently change the values of the control parameters,
- basing the analysis on average values of parameters carries the risk of misinterpretation,
- the simulation must be done in a timely manner to obtain the greatest benefit.

The simulation model design procedure includes the following stages [9]:

- identification of the simulated object using one of the two approaches: top-down, in which the main process is detailed into sub-processes and activities; bottom-up, which starts with defining all activities, and then grouping them into sub-processes and main processes,
- developing diagrams of the simulated process using IT tools (the number of hierarchy levels depends on the detail of the analyzed process),
- collecting input data and parameters, and then entering them into the simulation model,
- model verification, which comes down to comparing the behavior of the simulation model with the actual behavior of a given system (Figure 1).
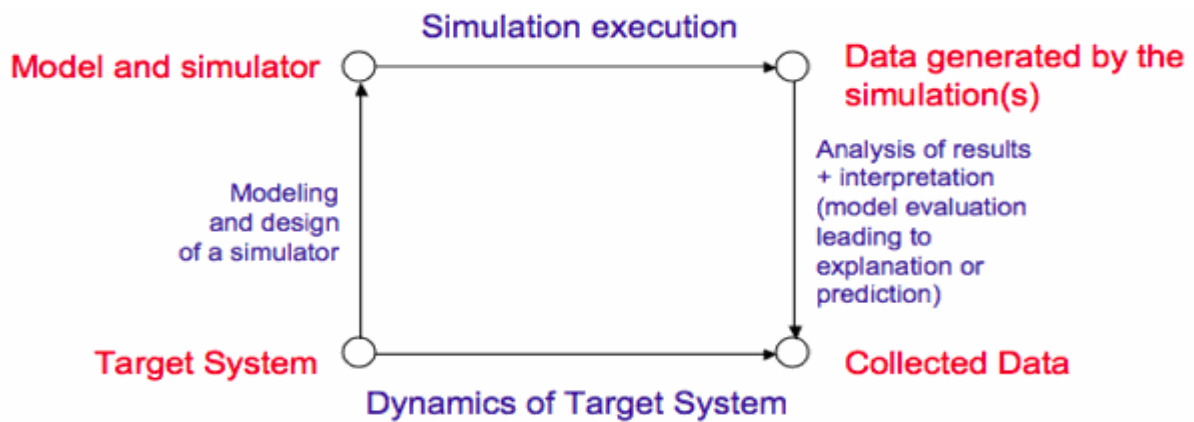

Figure 1. A general schema describing the usage of simulation as a predictive or explanatory instrument. Source: (Bandini, Manzoni, Vizzari,

Computer simulation also allows you to extend the operating time of the system, because it can be used to examine the detailed structure of changes that could not be observed in real time.

A Digital Twin is a mirror image of a physical process that is articulated alongside the process in question, usually matching exactly the operation of the physical process which takes place in real time (Batty, 2018). The term Digital Twin denotes a replica of a physical asset, process or system used for control and decision making [10].

Digital Twins are adopted by several disciplines. They have been applied to agriculture [8], Industry

4.0 in the context of smart manufacturing [5], prediction of the ergonomic performance in automotive industry Caputo, 2019.

A Digital Twin is expected to enhance city management and operations to achieve a smarter and sustainable city and a higher quality of life for its citizens. First implementations of Digital Twins in context of Smart Cities have arisen. In Zurich the city Digital Twin enhances city administration and support urban planning decision-making processes (Figure 2). To enable the use of the Digital Twin, open governmental data is being utilized in order to facilitate

contributions from the different stakeholders and their accessibility to the city data [11].
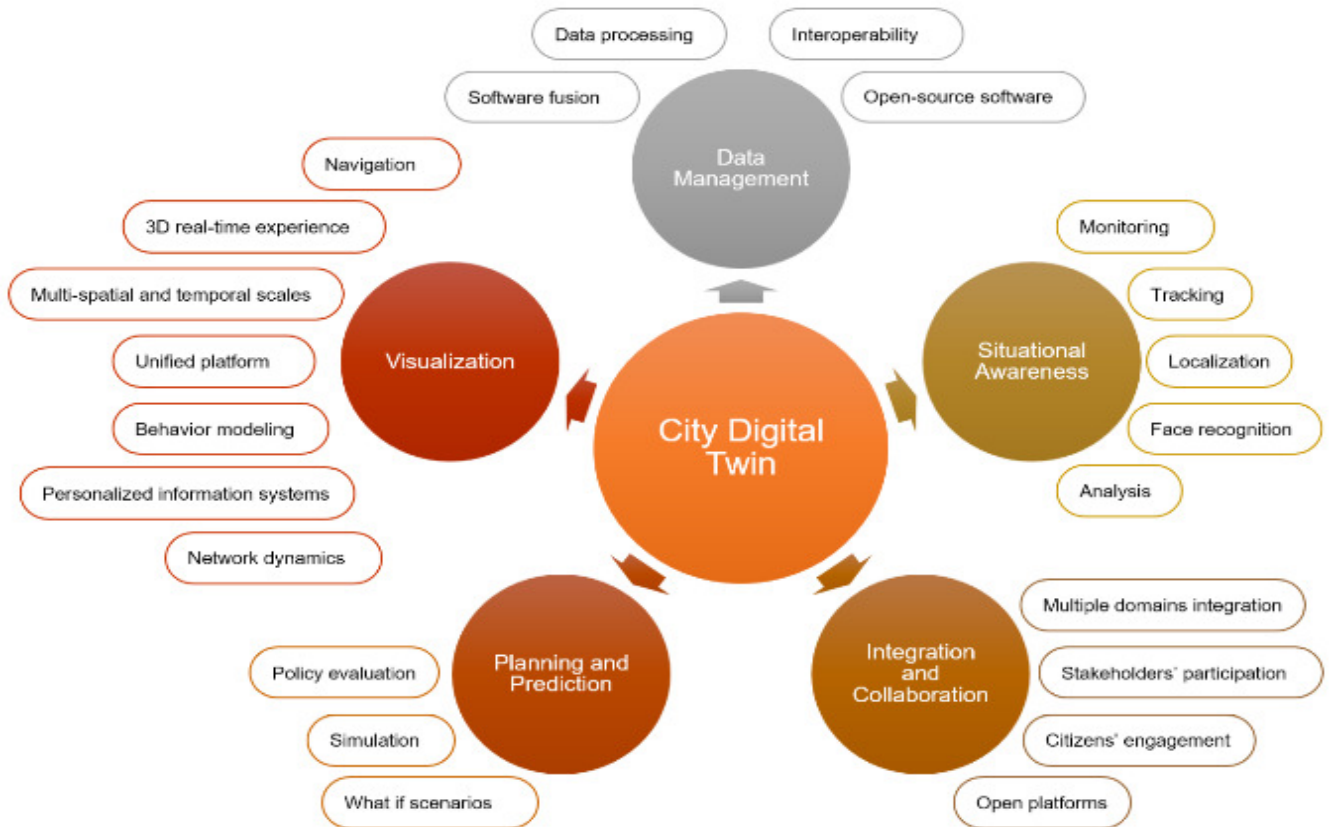

Figure 2. City Digital Twin potential. Source: [11]

Digital Twin is increasingly being explored as a means of improving the performance of physical entities through leveraging computational techniques, themselves enabled through the virtual counterpart [6]. Digital Twin very often starts life as a Digital Twin Prototype and helps in modelling, testing, optimization of a real product or asset. In its essence Digital Twin enables the application of a knowledgeable, data driven approach to the monitoring, management, and improvement of a product or a city asset throughout it's life-cycle. A digital twin can be perceived as an opportunity to enhance city planning and operability. Digital Twins enable performing simulations on the virtual model. Forecasting and optimization of the physical entity's performance are realizable, and thus the optimization of the physical counterpart's performance can be achieved. Digital Twins can also engage the citizens in creating new plans for the city and enhancing public decision-making.

Data describing real city processes is a key resource required to build and maintenance a relevant implementation of a Digital Twin that delivers reliable insights. Proper data availability is the one of most significant challenge in Smart City Digital Twin area. We can distinguish the following issues regarding data:

- lack of open data sources in specific cities referring to a selected city domains,
- large-sized, complex, and heterogeneous nature of the city data,
- lack of a widely accepted standards for the data models and design schemas to facilitate the development of the city models.

Open data is data that anyone can access, use and distribute. This definition, formulated by the Open Data Institute [12], can be applied both to public data (generated in the public sector, e.g. by government administration or other state institutions) and to research data. As with scientific publications, also with data there are technical and legal barriers that must be removed in order for data to be

considered open. Two stack models to assess the degree of data openness can be distinguished:

- The FAIR model (Findable, Accessible, Interoperable, Reusable) is a set of recommendations formulated by a group of experts from the FORCE11 organization, which should guide people opening research data [13]. This model identifies four most important aspects of open data: it should be well searchable, accessible, interoperable and reusable. The authors strongly emphasize that due to the way data is used in the modern world and in modern science in particular - data should be available both in a human-readable form and in a form suitable for machine analysis.
  - o The 5-star open data model was developed primarily with public open data in mind, but can be applied to research data [14]. In this model, the removal of legal barriers means that the data is awarded one star, the next four stars relate to the removal of technical barriers to the use of data.

Open (government) data is information collected, provided, or paid for by public authorities (also known as public sector information) that is made freely available and re-used for any purpose.

Problems using open data:

- Purpose - open data allows you to look deeper into a specific topic that we want to know more about. Economic operators can also use open data to refine their customers' profiles and better adapt to their needs. Whether used for private or commercial use, open data offers many possibilities.
- License - the license allows the use of data in a way that interests us (eg if we create a commercial application, is it allowed to re-use the data in a commercial way). The license may require you to identify the data publisher when used, i.e. we must provide the data owner when we make the product or service available. This requirement is called attribution.
- Data format - when we find that a specific data set contains exactly what we need, we can download it in one of the available formats. Based on our IT knowledge, we select the most appropriate type of file. The most common format for tabular data is ".csv". It allows you to add information to a file and perform calculations using the data contained in it. Data sets that can be changed are published in an open format. Most datasets are available in an open format, but please note that some formats (e.g. ".pdf") are not changeable.
- Data quality - the page from which we want to download the data set should contain the date of the last modification of the file. If we need data from a

specific period, it is necessary to check whether the information about the time period is provided or whether the file has been recently updated. You should also make sure that the information you expect to find in the file is actually included in it and that you recognize individual labels.

Checklist developed by the Open Data Institute [15]:

- form:
  - o how is the data processed?
  - o are they in a processed or unprocessed form?
  - o how will the form affect the analysis / product / application?
  - o what syntactic (language) and semantic (meaning) transformations will be required?
  - o are they compatible with other, already owned data sets?
- quality:
  - o how up-to-date is the data?
  - o how regularly are they updated?
  - o are all fields and data context understandable?
  - o how long will they be published? what is the publisher's commitment?
  - o what do we know about data accuracy?
  - o how is the missing data problem solved?

Open public data is the data of institutions and offices that anyone can use. On the basis of open public data, more and more modern products and services are created in Europe and around the world. Open data is a real source of real savings in money and time for administrations and citizens. Citizens, including entrepreneurs, can use public data resources to pursue their own goals, developing their business or research.

One of the conditions for digital development is quick and effective access to high-quality data, which allows for the creation of more innovative solutions, e.g. in the area of the so-called artificial intelligence, or to put it more precisely, automation and prediction. The European Data Strategy [14], which was developed by the European Commission, identifies the openness of high-quality data and the value of data as one of the pillars of building the competitiveness of the European Union economy. Therefore, in 2021, the adoption of implementing acts was planned, which will enable the public sector to make data sets available in a machine-readable format and via application programming interfaces (APIs). Data openness is indicated as a key element to stimulate innovation in many sectors of the economy, but also in science, and machine learning technologies, natural language processing or the Internet of Things require increasing the supply of the said data [16].

III. NEW PARKING REQUIREMENTS

Environmental friendly transport system is one of the most significant parameters of the smart city. There is some kind of the tacit contradiction between dynamic (car movement) and static (car parking) aspects of the traffic in contemporary European cities, which are more and more closed for private and business transport. In the European Union won the static concept, which means that the big transport vehicles cannot enter city closed system; they must park on the city boarder. Also private car movement is strongly oriented toward city parking system. The real transport is more oriented toward city outside environment; in the city it is totally restricted to minimum with especially for gasoline vehicles.

The EU ecology strategy and environmental regulations are mainly oriented toward:

- significant reduction of CO2 emission,
- development of environmental friendly energy sources,
- enhancement of electric and hybrid car production.

One of the real determinant for such EU development is obviously actual geopolitical situation connected with diversification of the global energy sources. These, above mentioned principles, create more sophisticated expectations for:

- car industry,
- transport system within EU and
- smart city models.

On the base of EU ecology model we can identify the following new smart city problems:

- reduction to minimum the quantity of gasoline vehicles within cities,
- supporting and extension of the city routes fort electric bikes, scooters, motorbikes, autos and small trucks,
- renovation of existing and construction of the new parking with electric loading systems.

We notice in the last years significant grow of the global electric cars sales volume especially of the following car companies:

- Toyota,
- KIA,
- Tesla and
- European producers.

This trend is inevitable; e.g. Volkswagen and other German car manufacturers invest strongly and construct new auto electric battery factories. The lack of electric battery loading stations creates another, big problem for electric car users. Therefore the authors of this paper suggest analyses and optimize of city parkings in the context of:

- reconstruction of the existing parkings toward some kind of "electric parking plant", where the car drivers can load their electric vehicles,
- construction of the new parkings, fully equipped in the modern electric loading systems.

Such investments will give for parking owners significant competitive advantage.

IV. OPEN DATA IN SIMULATION OF PARKING SPACES ON THE EXAMPLE OF WROCŁAW

RESARCH METHOLOGY

The aim of this research is to find simply measure of character of parking traffic. To achieve the aim was using data gathered from Wroclaw city portal.

The data for the study come from the website of the city of Wrocław, https://www.wroclaw.pl/open-data/dataset/zapelnienieparkingowodczytza48h_data. The data is open data and collected from 4 car parks in Wrocław, Hala Stulecia, National Forum of Music, Nowy Targ and St. Anthony. St. Anthony is the subject of our research. The data is published on the website every 48 hours, which means adding more data.

Table 1. Structure of data from the website of the city of Wroclaw

_id	Registration time	Number of free parking spaces	Number of v. entering	Number of v. leaving	Name of car park
252	05.06.2022 00:00:00	175	2	2	Nowy Targ
2	05.06.2022 00:00:01	787	0	0	Parking Hala Stulecia
1	05.06.2022 00:00:02	41	0	2	ul. św. Antoniego
200	05.06.2022 00:05:00	176	2	2	Nowy Targ
4	05.06.2022 00:05:01	41	1	1	ul. św. Antoniego
3	05.06.2022 00:05:01	787	0	0	Parking Hala Stulecia

Shared public data is composed of six columns. The structure od the file is presented in **Błąd! Nie można odnaleźć źródła odwołania.**3.

The indicator of a type of parking traffic was defined as a deference between number of cars income into parking and number of cars goes outside of the parking in the same period of time

Type traffic=number of enters cars - number of leaving cars

That defined indicator allows for the study of the character of parking traffic. When Type traffic is positive that means, that entering traffic is higher than leaving traffic and the number of parking spaces reduce. In case when Type traffic is negative that means, that entering traffic is lower than leaving traffic and the number of parking space increase. The sign of Type traffic determines the direction of parking traffic in the time period, while the value of Type traffic talks about the intensity of the direction of traffic.

That simply measure could be useful especially with combination with the data visualization technique.

To visualize the Type traffic, first of all, the original parking data must be transformed into new structure. The model of data show **Błąd! Nie można odnaleźć źródła odwołania.**2. To transform the model of parking data was use Python programming language and the Pandas package.

Table 2. Transformed structure of the dataset.

Column's name	Description
Id	Record id
Id parking	A number representing the car park name
Register time	Time of register the event. The time format is hour:minutes
Day number	A number of the day of the week
Number of entering cars	A number of vehicles entering the parking. Value is a sum of enter cars in period of 15 minutes
Number of leaving cars	A number of vehicles leaving the parking. Value is a sum of leaves cars in a period of 15 minutes

It seems that the visualization of the Type traffic indicator in 15th minutes periods is too high, and this is the reason why records were aggregated to one hour period. The new model of data is presented in Table 3.

Table 3. New format parking data with Type traffic indicator

Column's name	Description
Id	Record id
Id parking	A number representing the car park name
Date	Date of register the event. The format is year-mont-day
Day number	A number of the day of the week
Type traffic	Difference between the number of input cars and the number of output cars in one hour period

1. Files with data were downloaded every 2 days and stored in shared drive to collect data covering 2 days of car park functioning.
2. Rows were merged to generate 1 consistent data set presenting car park functioning through 2 days.
3. Data was imported to Microsoft Power BI Desktop.
4. Preparation of data visualisations in Microsoft Power BI Desktop to identify possible car park deficiencies.
5. Transformation of original file int a structure enabling a comparative analysis between events of every day. The structure is presented in Table 4.
6. Data analysis in Python language programming with Pandas package.
7. Preparation of data visualisations in Python language with Plotly package to identify possible car park deficiencie

Table 4. Transformed structure of the dataset.

Column's name	Description
Id	Record id
Id parking	A number representing the car park name
Register time	Time of register the event. The time format is hour:minutes

Day number	A number of the day of the week
Number of entering cars	A number of vehicles entering the parking. Value is a sum of enter cars in period of 15 minutes
Number of leaving cars	A number of vehicles leaving the parking. Value is a sum of leaves cars in a period of 15 minutes

RESULT

Figure 3 shows the number of vehicles entering the parking lot from 4th of April to 7th of April, and Figure 5 shows the number of vehicles leaving the parking lot.
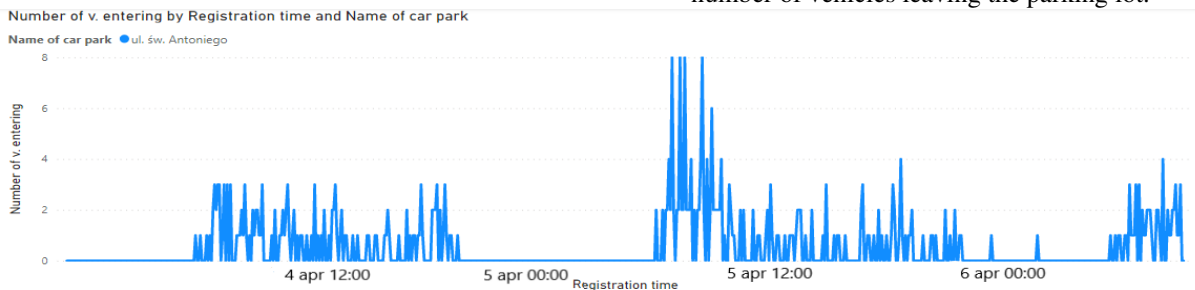

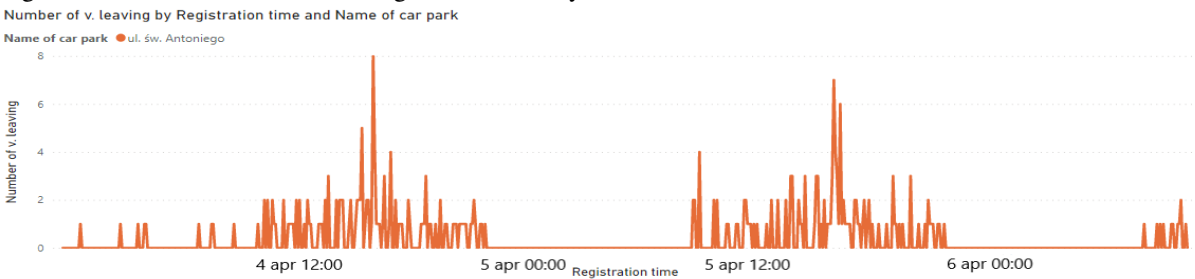Figure 3. Number of vehicles entering the St. Anthony

Figure 4. Number of vehicles leaving the St. Anthony

A sum of the number of entering cars and the sum of the number of leaving cars in an hour period should let for better understanding visualization of the Type traffic indicator.
To visualise date Plotly package was used. In the Figure is presented visualization of the Type traffic in one hour period. Thera are data comes from four days from 4th of April to 7th April

Figure 5 is presented a visualization of the Type traffic in a 24hour period from 0 a.m to 23 p.m. The X-axis represents one-hour time periods. Every period shows an aggregated value of Type traffic. There are data comes from four days from the 4th of April to the 7th of April.
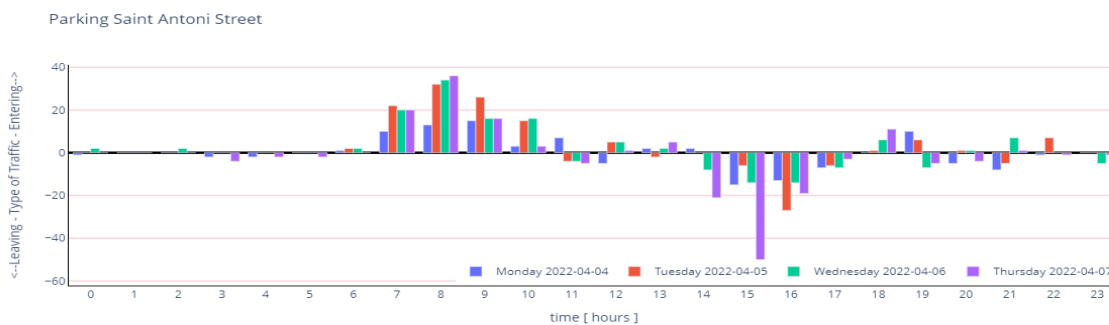

Figure 5. Visualization of the Type of traffic

Analysing the graph below it is clear, that Type traffic in period from 6 am to 11 am has a positive value, which means entering traffic is higher than leaving traffic, and free parking spaces reduce. The character of parking traffic is changing from 2 pm to 6 pm, where Type traffic is negative, which means parking spaces increase. The same character of parking traffic repeating in every day.
The Type traffic indicator could be a useful measure to describe the character of parking traffic.

V. DISCUSSION

The aim of this paper was to highlight the possibilities offered by intelligent IT tools in the area of simulating selected processes occurring in urban space.

Authors decided to indicate possibilities of using Digital Twin method in analysis, simulation and support of intelligent solutions related to the management of parking space. Due to the pilot character of the presented research, data obtained from open sources made available by Wroclaw were used in the analysis.

At this stage of designed research, the authors diagnosed a significant problem in gaining access to data from different cities in Poland. In the initially defined research procedure, it was planned to contrast data on functioning of parking lots from selected, provincial cities in Poland. Unfortunately, except for Wroclaw, the authors were not able to obtain credible, reliable and up-to-date data on occupancy of urban parking lots in other voivodship cities. It should be pointed out that especially the parameter of actuality and systematic data refreshing is important for the credibility of created simulations. That is why, the authors decided to present a study using data made available by Wroclaw, which are updated every 10 minutes and concern 4 city parking lots.

The conducted research has shown that the availability of open parking lots allows you to conduct an analysis in terms of the occupancy of parking spaces, as well as to observe various types of anomalies related to it. What they lack is a link to other data to identify this fact. However, it is possible through additional data acquisition, e.g., about events taking place in the vicinity of a given car park or in the city. These analyzes will allow you to build a specific model, as well as determine the conditions for its functioning. As an example, we can point to, for example, the simulation of store queues, which can be used to determine the potential behavior of customers in the plotted conditions. In the field of parking space management, it will be possible to determine the occupancy of parking spaces, as well as, to a further extent:

- creating additional applications informing which parking space is free,
- which seats will be occupied at a certain point in time,
- total occupancy of the car park.

In addition, if additional metadata such as location, city, nearby public facilities are used, it may be possible to accurately plan and build additional/new parking spaces.

VI. CONCLUSION

Both visual analyses show us, that the transformation original dataset allows us to better understand parking traf-

fic. The moments of intensive traffic are presented. In such moments traffic related issues occur, that Digital Twin and simulation can address. We can observe e.g the repeating distribution of the number of entering cars and also unexpected traffic caused by external factors like public events. Shared datasets by the Wroclaw are good datasets to start the process of parking traffic analysis. Having such files it is possible to start building a simulation addressing the issue.

REFERENCES

[1] Bandini, S., Manzoni, S., Vizzari, G. (2009). Agent Based Modeling and Simulation: An Informatics Perspective, Journal of Artificial Societies and Social Simulation 12 (4).

[2] Batty, M. (2018). Digital twins. Environment and Planning B: Urban Analytics and City Science, 45(5), 817-820.

[3] Caputo, F., Greco, A., Fera, M., & Macchiaroli, R. (2019). Digital twins to enhance the integration of ergonomics in the workplace design. International Journal of Industrial Ergonomics, 71, 20-31.

[4] Dullinger, K.H. (2009). Simulation in der Logistik - neue Anwendungsfelder. LogForum, Vol. 5, Issue 3, 1-12.

[5] Jiang, Y., Yin, S., Li, K., Luo, H., & Kaynak, O. (2021). Industrial applications of digital twins. Philosophical Transactions of the Royal Society A, 379(2207), 20200360.

[6] Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the Digital Twin: A systematic literature review. CIRP Journal of Manufacturing Science and Technology, 29, 36-52.

[7] Koliński, A., Śliwczyński, B., Golińska-Dawson, P. (2019). Wykorzystanie symulacji jako narzędzia wspomagającego proces oceny efektywności produkcji w przedsiębiorstwach produkcyjnych. E-mentor nr 3 (75) / 2018.

[8] Pylianidis, C., Osinga, S., & Athanasiadis, I. N. (2021). Introducing digital twins to agriculture. Computers and Electronics in Agriculture, 184, 105942.

[9] Rodawski, B. (2006). Simulation of logistics processes. LogForum, Vol. 2, Issue 1, 1-15.

[10] Vatn, J. (2018, June). Industry 4.0 and real-time synchronization of operation and maintenance. In in Proceedings of the 28th International European Safety and Reliability Conference (pp. 681-5).

[11] Shahat, E., Hyun, C. T., & Yeom, C. (2021). City digital twin potentials: A review and research agenda. Sustainability, 13(6), 3386.

[12] www1   https://theodi.org/what-is-open-data   (date   of   access: 15.04.2022).

[13] www2 https://www.force11.org/group/fairgroup/fairprinciples (date of access: 17.04.2022).

[14] www3 https://www.parp.gov.pl/component/content/article/71119:otwarte-dane-filarem-innowacyjnej-gospodarki-co-przyniesie-nowa-ustawa#_edn1 (date of access: 15.04.2022).

[15] www4   http://theodi.org/guides/the-open-data-consumers-checklist (date of access: 17.04.2022).

[16] www5   https://data.europa.eu/pl/trening/what-open-data   (date   of access: 15.04.2022).