RESEARCH ARTICLE

Exploring city digital twins as policy tools: A task-based approach to generating synthetic data on urban mobility

Gleb Papyshev[1] (ID) and Masaru Yarime[1,2,3,*] (ID)

[1]Division of Public Policy, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong
[2]Department of Science, Technology, Engineering and Public Policy, University College London, London, United Kingdom
[3]Graduate School of Public Policy, The University of Tokyo, Tokyo, Japan
*Corresponding author. E-mail: yarime@ust.hk

Abstract

This article discusses the technology of city digital twins (CDTs) and its potential applications in the policymaking context. The article analyzes the history of the development of the concept of digital twins and how it is now being adopted on a city-scale. One of the most advanced projects in the field—Virtual Singapore—is discussed in detail to determine the scope of its potential domains of application and highlight challenges associated with it. Concerns related to data privacy, availability, and its applicability for predictive simulations are analyzed, and potential usage of synthetic data is proposed as a way to address these challenges. The authors argue that despite the abundance of urban data, the historical data are not always applicable for predictions about the events for which there does not exist any data, as well as discuss the potential privacy challenges of the usage of micro-level individual mobility data in CDTs. A task-based approach to urban mobility data generation is proposed in the last section of the article. This approach suggests that city authorities can establish services responsible for asking people to conduct certain activities in an urban environment in order to create data for possible policy interventions for which there does not exist useful historical data. This approach can help in addressing the challenges associated with the availability of data without raising privacy concerns, as the data generated through this approach will not represent any real individual in society.

Policy Significance Statement

The significance of this article for policymakers is shown in three aspects. First, it provides an overview of the challenges associated with privacy, availability, and applicability for predictive simulations of the data that are being used for city digital twins. Second, it discusses how the application of synthetic data can address these challenges based on the examples from other domains of application. Third, it proposes a new methodology for urban mobility data generation, which can address the aforementioned challenges and serve as a tool for citizen engagement.

---

1. Introduction

The growing complexity of the world requires new analytical tools to be deployed for harnessing the potential benefits of evidence-based policy. City digital twins (CDTs) create opportunities for city

officials to embrace the notion of simulation governance and expand the reach of contemporary planning techniques. The simulative notion of the digital twin (DT) allows it to be turned into something more than a digital copy of the city, as the virtual nature of this approach to modeling provides the necessary infrastructure for a wider spectrum of imaginative ideas to be applied to the design of smart cities.

However, while the spectrum of possible applications of this technology is vast, especially in the context of policymaking where numerous scenarios can be tested in a virtual environment of a DT before the actual policy intervention is conducted, the empirical research on the applications of this technological tool to policy domains is scarce. There are some prominent examples of empirical research in this field (Nochta et al., 2021), but overall, the challenges of utilizing CDTs in the process of policymaking from a less engineering-oriented perspective are rarely discussed in the academic literature.

This article discusses the technology of CDTs and their potential utilization by policymakers. The article also provides some critical remarks which need to be taken into account when dealing with CDTs, as well as proposes a methodology for data generation, which can be applicable for CDTs. The article proceeds as follows.

The first part of the article discusses the history and conceptual underpinnings of this technology and analyzes the capabilities of one of the most advanced projects in the field—Virtual Singapore. The second part discusses the limitations of the current generation of CDTs. Concerns related to data privacy, availability, and their applicability for predictive simulations are discussed (Bektas and Schumann, 2019; Kieu et al., 2020) alongside the role that synthetic data can play in addressing these concerns (Nikolenko, 2019). The final section of the article proposes an alternative task-based approach to urban mobility data generation. This approach is informed by the practices of data labeling, urban data games, and games with a purpose. Under this proposal, the city authorities can establish services responsible for asking people to conduct certain activities in an urban environment in order to create data for possible policy interventions for which there does not exist useful historical data. This can potentially allow the governments to collect unique data about possible behavioral responses to certain interventions, while simultaneously not violating any privacy concerns as the data generated through this approach will not be representative of any single individual.

2. CDTs of Current Generation

2.1. DT concept

A DT is a virtual representation of the characteristics and behaviors of a physical object. The purpose of the creation of a DT is to model and predict the lifecycle of a system (Jones et al., 2020). The concept of DT originated in the works of Michael Grieves and John Vickers at NASA in 2003 (Grieves and Vickers, 2017). It was not very specific at the time with only three key characteristics of the concept being articulated: physical product, virtual product, and connections between them (Tao et al., 2019; Jones et al., 2020).

However, the practice of creating a duplicate of a system has been explored and practised at NASA even earlier starting from the 1960s, when the so-called "first digital twin" (Boschert and Rosen, 2016; Enders and Hoßbach, 2019) was created to simulate the conditions on board of Apollo 13 in real-time. It was the first example when a duplicate of a real system was used to mimic the system's conditions in real-time to avoid the failure of the mission (Ferguson, 2020). In 2012, the concept of DT was redefined at NASA to be understood as an integrated multiphysics, multiscale, probabilistic simulation of a system, which mirrors the life of a corresponding physical artifact based on historical data, physical model, and real-time sensing (Glaessgen and Stargel, 2012).

The concept of DT has become a popular research topic (Tao et al., 2019) with different approaches to the issue. Tao and Zhang (2017) propose to expand the three-dimensional model of a DT comprising of a physical part, virtual part, and connections between them with data and service. Despite the debates about the definition of DTs, the core concept behind it is a system that couples physical and virtual entities of a system to leverage the benefits from both (Jones et al., 2020). The main feature of this approach is the

ability to predict the future state of a system when tested under different conditions based on data-driven analytics and simulations in real or hypothetical conditions (Cioara et al., 2021).

According to Grieves and Vickers (2017), if applying the model of DTs, then the creation of a physical object starts in the virtual space through the creation of a DT prototype. Once the modeling and simulations of a system are conducted and potential obstacles are understood, the physical object is being built, and the data connections between the virtual and the physical domains are established. DTs enable control over physical entities via digital objects without human intervention (Enders and Hoßbach, 2019). This is a unique feature of this approach, which is different from a digital shadow—when a change in the physical object causes a change in the virtual object, but not vice versa (Kritzinger et al., 2018).

DTs have gained popularity in the manufacturing literature (Malik, 2021; Rožanec et al., 2021) due to their ability for simulation, which is argued by some researchers to be the key feature of this approach (Kuts et al., 2019). DTs are also widely used in the aerospace engineering industry (Ríos et al., 2015) because they can replicate extreme conditions which cannot be physically performed in the laboratory setting (Sharma et al., 2020). A similar approach is exercised in the automotive industry, where DTs which comprise the whole car, its mechanics, electrics, software, and physical behavior are used to simulate most steps in the product's lifecycle in a virtual environment without the need to build numerous physical prototypes (Dharmani and Lulla, 2020). After the tests are over, the physical prototype is built and the connections between the model and the physical object are established. Other projects include the creation of a behavioral DT of a driver, through which warning and instructions about safer driving are provided in real-time (Chen et al., 2018).

With the advancements of this approach, it is now applied to the modeling of more complex systems. There are attempts to apply DTs in a smart energy grid; however, replicating the behavior of a physical energy asset is still problematic (Cioara et al., 2021). This approach is also being applied to socio-technical systems such as manufacturing plants (Park et al., 2019), where, for example, the creation of the DT for the whole plant can be the next step in the development of the DT for the artifact (such as a vehicle in the automotive industry) (Biesinger and Weyrich, 2019). Simulations allow for making insights into complex manufacturing systems in the virtual domain before translating them into the real world (Mourtzis, 2020). Computer-rendered avatars of humans are also used in these simulations in order to test their ergonomic viability (Malik, 2021). Livestock farming is another domain where a proposal for the utilization of DTs has been made, as the installation of sensors for measuring the environmental factors in real-time can be applied for simulations based on which decisions about the optimal CO_2 and temperature levels can be made in the near-real-time (Jo et al., 2018). This approach has been applied to even more complex socio-technical systems such as cities—arguably one of the most complex artifacts created by humans.

2.2. City DTs

The growth of computing capabilities, ubiquitous data flows and general interest in the application of DTs has expanded the application of this methodology from modeling of physical objects to the modeling of complex socio-technical systems. As this approach has gained prominence, DTs of smart cities are now created in order to analyze the data about urban complexities across time and scale (Francisco et al., 2020).

CDTs are a virtual replica of a city, which is continuously informed by the processes that take place in a real city via real-time data connections (Mohammadi and Taylor, 2017). CDTs can also be realized as systems of interconnected twins created on the scale of a block or a district, with a large number of observations interacting in the mathematical model—due to the volume and unknown relationships between the data, machine learning approaches can be useful for simulations (Ruohomäki et al., 2018). Pilot projects of smart CDTs are being developed in numerous cities across the globe, including Singapore, Glasgow, Helsinki, and Boston, with Helsinki expanding this notion to provide virtual tourism services with the usage of Virtual Reality technologies. The approach introduced by Grieves and Vickers (2017) is being utilized in a newly built Indian city, Amaravati, which is the first attempt to build a city

starting from building its DT—thus, allowing everything to be modeled and simulated before embarking on building it (smartcitiesworld.com, 2020).

The academic discussion about CDTs has also been vibrant recently, which can be attributed to the growing urbanization rates and the rise of the Internet of Things (IoT) and data analytics (Fuller et al., 2020), as well as the need for the introduction of digital services to the city management (Soe, 2017). The creation of CDTs is dependent on the availability of massive urban data generated every day via different sensors installed in the city and technical foundation, such as IoT and 5G (Deren et al., 2021). Deng et al. (2021) argue that the creation of DTs is an inevitable goal of digital transformation of a city, which is countered by Nochta et al. (2021) who argue that it is too early to argue that CDTs bring a paradigm shift to urban modeling.

Deren et al. (2021) argue that the four major characteristics of CDTs include accurate mapping (modeling of the physical aspects of the city), virtual–real interaction (aspects of the physical environment can be observed in a virtual environment), software definition (simulations are conducted on software platforms), and intelligent feedback (early warning of potential adverse effects and potential dangers). A systematic analysis of the CDT literature reveals that the potential for its applications can be broadly divided into five themes: data management, visualization, situational awareness, planning and prediction, and integration and collaboration (Shahat et al., 2021).

In this typology, data management becomes the crucial pillar for the development of the CDT, as the ability to manage and process data coming from different sources in the city is dependent on the ability to integrate this data through various data standardization and data sharing programs (Ruohomäki et al., 2018). The visualization of the city processes is another key feature of CDTs, where visualization of social processes is among the key challenges (Shahat et al., 2021). The situational analysis allows for the monitoring and analysis of the activities taking place in the city, which include monitoring of the health of the citizens based on the data from personal health devices (Laamarti et al., 2020), monitor daily building energy consumption (Francisco et al., 2020), or detect motion (Dou et al., 2020). There are also experiments about how to use CDTs in the process of planning future city operations scenarios, for example, traffic behavior in the city under different conditions (Dembski et al., 2019). Integration of different aspects of CDTs in one platform presents a challenge in the field (Shahat et al., 2021), with some proposals arguing that the development of several DT models for one city can be more feasible than developing a single model to capture all the complexities of the city's operations (Wan et al., 2019). Other approaches include a combination of DTs of different scale built for various purposes with different approaches for capturing the complexities of a nation's infrastructure (Lu et al., 2020).

The application of crowdsourced visual data for the estimation of the geospatial information of vulnerable objects in the cities can be integrated with the 3D virtual city model for immersive visualization, which can be used for making more informed decisions about infrastructure management and run what-if simulations for disaster situations (Ham and Kim, 2020). CDTs can be used as a tool to run simulations for different disaster management scenarios, where based on the data collected from the sensors installed in the city a realistic reaction of a physical space to a certain natural disaster can be simulated, which can be combined with natural language processing-enabled system for monitoring of the activities on social media (Dogan et al., 2021). However, in times of urgent crises such as the Covid-19 pandemic, high-quality long-term data crucial for CDTs is unavailable. A collaborative CDT model based on a federated learning methodology, when different CDTs learn a shared model while keeping all the training data locally, can be leveraged in situations of emergency (Pang et al., 2020). A model for the incorporation of the mental features of decision-making in CDTs has been proposed (Klebanov et al., 2019), as well as a proposal for integration of artificial intelligence (AI) algorithms for situation assessment in emergency situations for disaster management (Fan et al., 2021).

The technological approach utilized in the CDTs allows for the integration of real-time data into a 3D model of a city (Shahat et al., 2021) which can be used for the analysis of various city domains (Dou et al., 2020). The integration of the real-time data flows into the 3D can be used for tracking the widespread of information and determine vulnerable objects during disasters (Ham and Kim, 2020; Fan et al., 2021). This type of information can inform city officials about what areas to evacuate first (White et al., 2021).

Another domain where real-time data inflow can be utilized is traffic control and congestion pricing, though this approach is prone to the privacy critique (Bliss, 2019). Near real-time efficiency of a building in terms of energy consumption can be used to determine variations from conventional benchmarks for real-time energy management (Francisco et al., 2020). Certain difficulties still exist in this domain, however, as the ability of a CDT to be updated in near real-time from the received information from the physical object has been proven, but the inverse connection and data transfer from the virtual to the physical are still challenging (Shahat et al., 2021).

Although CDTs are positioned to grow from experimental playgrounds for city planners (Hemetsberger, 2020) to an effective tool for city management (Wong et al., 2020), currently, CDTs only abstract a small fraction of the processes that shape the way in which the social and economic functions of the city work (Batty, 2018). Nochta et al. (2021) argue that CDTs can be useful at the early stages of policymaking, where they can be utilized for identifying inconsistencies between sectorial policies, supporting interdisciplinary policy design, and improving the efficiency and effectiveness of modeling. Other applications of CDTs in policymaking support include simulations of traffic scenarios (Dembski et al., 2019), identification of vulnerable physical objects in the city under "what if" simulations of disasters (Ham and Kim, 2020), and measuring effects of potential urban planning interventions on climate (Schrotter and Hürzeler, 2020).

2.3. Virtual Singapore

Virtual Singapore serves as a good example in the discussion about CDTs, as it is one of the most advanced projects in the field (Geddie and Aravindan, 2018). The project was granted $73 million by the National Research Foundation of Singapore and was developed by the French software company Dassault Systemes. It is a dynamic 3D model of the whole city of Singapore as well as the necessary technical infrastructure to turn it into a collaborative data platform (National Research Foundation Singapore, 2018). Other stakeholders involved in the project include Singapore Land Authority and Government Technology Agency (Nativi et al., 2020), with the former providing static data about the aboveground structures in the city (Guerrini, 2016).

The static data from the governmental agencies are not limited to the 3D models of the aboveground structures in the city. They also capture detailed information about the built environment of the city up to the scale of the building, which includes information about its geometry and materials it is made of (Koçer, 2020), as well as the data about the flora of the island (Gobeawan et al., 2018). The data ecosystem also includes real-time data which is being obtained via IoT sensors deployed in the city (Guerrini, 2016). The data generated by other city-scale platforms, such as OpenMap Singapore, are also added in the DT model of the city (National Research Foundation Singapore, 2018). The CDT has recently been updated by the data about underground structures of the city, with this part of the project being titled "Digital Underground" (Yan et al., 2019, 2021). With the addition of the data about the belowground systems of the city, the CDT of Singapore now includes 20 core datasets: 3D airspace, vegetation, 3D road, 3D building models, cadastre, land use, administrative bodies, waterbody, geodetic control, orthophoto, 3D coastline, digital terrain model, point cloud, 3D reality mesh, imagery, positioning infrastructure, 3D address, building information modeling (BIM), 3D underground asset, and 3D geology (Schrotter, 2020).

Being a 3D model of the city, in comparison to the usual 2D models, Virtual Singapore is not only able to incorporate more data layers in the model, including the data about water bodies, vegetation, and transportation infrastructure, but also capable of showing the information about the curbs, stairs, or the steepness of the hill in the city (Dassault Systemes, 2018). Possible use cases of such information are identification of barrier-free routes for the elderly and disabled people, sharing of the information about wild animals in the city, tracking elderly people who have dementia, and detecting frequently used paths and spots in the city (Singapore Land Authority, 2014).

The applications of the Virtual Singapore model can include the visualization of the 3G/4G internet coverage areas in the city in order to determine which parts of the city suffer from poor connection; agent-based simulations of crowd dispersion, pedestrian movements, and transport flows; collaborative design

of the pathways around new amenities in the city—the way such pathways will be built will influence the pedestrian flows, which can be simulated in the DT environment to determine the right intervention; and examination of the most efficient way of installing solar panels on the roofs to achieve higher energy production (National Research Foundation Singapore, 2018).

The project positions itself as a collaborative tool where different stakeholders can potentially conduct virtual experiments in the urban environment (National Research Foundation Singapore, 2018). However, some officials consider the system to be too dangerous to allow everyone to experiment with it, because, for example, militants or terrorists can use the information about the height, location, and the view from the buildings to plan their attacks (Geddie and Aravindan, 2018). Due to these reasons, Virtual Singapore will not be connected to the worldwide web, and the model has not yet been made publicly available, so the citizens are unable to interact with it (Geddie and Aravindan, 2018; White et al., 2021).

A common critique for the CDTs like Virtual Singapore and the simulations that can be conducted in these environments is their overreliance on historical data, which limits its predictive capabilities for the extreme situations for which there are no data available (GeoTwin, 2020)—which is the case for the Virtual Singapore project (Holstein, 2015). Other critical takes on the current generation of CDTs include lack of mechanisms for citizen engagement, interaction, and feedback report, as well as underutilization of urban mobility data (White et al., 2021). Concerns related to the privacy protection of the micro-level data about individuals used for agent-based simulations in CDTs are also voiced (Bektas and Schumann, 2019; Bliss, 2019).

3. Shortcomings of the Current Generation of DTs

3.1. Data for DTs

With the digitalization of cities, different types of data can be generated within the urban boundaries (White et al., 2021), providing detailed information about transportation (Menouar et al., 2017), water supply (Parra et al., 2015), waste management (Medvedev et al., 2015), and power generation (Oldenbroek et al., 2017). However, meaningful knowledge discovery from such distinctively different data remains problematic due to issues of noise and missing values in the data or linking structured and unstructured data (Mohammadi and Taylor, 2020). The nonexistence of certain types of data (Hemetsberger, 2020), the lack of data standard, and the unwillingness to share the data by different institutions are also challenges that may slow down the process of the development of CDTs (Wong et al., 2020).

According to Wong et al. (2020), the data that can be integrated into the CDT model can be represented as a 2×2 matrix, with the 2D/3D and Dynamic/Static dichotomies. In such a model, 2D dynamic data include navigation, remote facility monitoring, pandemic management and tracking, crowd and traffic control, climate diagram, and city dashboards. The 2D static data include maps, administrative boundary, and building plans. 3D dynamic data include navigation (3D), building management system, maintenance of underground pipes, microclimate and airflow analysis, command and control center/crisis management, whereas 3D static data include building information modeling, geographic information system, and design visualization (interior/venue).

Fuller et al. (2020) argue that the number and quality of connection of IoT devices installed in the city are crucial for gathering enough relevant data for CDTs. Although feeding a lot of different data (Big Data) into the DT and then applying machine learning techniques for making predictions is considered to be a viable way of utilizing the power of IoT in the DT (Qi and Tao, 2018), some researchers question these techniques because some level of verification of the validity of the prediction is required, which is not always possible (Boje et al., 2020).

The high demand for city data from CDTs raises concerns about the safety and privacy of the data because certain applications such as real-time traffic flow simulations would require an excessive amount of individualized information (Bliss, 2019). Although not all CDTs of the current generation include urban mobility data (White et al., 2021), the potential issues related to data aggregation infrastructure on the city-level pose significant concerns among various stakeholders, affecting the trust levels in the technology (Ismagilova et al., 2020).

3.2. Human aspects of CDTs

The digitalization of city services and ubiquitous computing provides numerous channels through which up-to-date human mobility data can be aggregated at various temporal and spatial scales (Luca et al., 2020). The ways in which mobility data can be aggregated include GPS trackers embedded in mobile phones (Zheng et al., 2010) or vehicles (Pappalardo et al., 2013), the connection of the mobile phone to the cellular network (González et al., 2008), and geo-tagged posts on social media (Blanford et al., 2015).

There are numerous research projects that focus on the mining of the human mobility trajectory data (Jiang et al., 2013) and finding statistical patterns in it (Barbosa-Filho et al., 2017). An abundance of this type of data allows for the application of new techniques to its analysis in order to predict the next location an individual will visit based on the historical data (Burbey and Martin, 2012; Wu et al., 2018) or forecast the flows of people on a geographic region (Ebrahimpour et al., 2019). The applications of these methodologies for policymaking are vast and include potential public emergency detection, public safety and traffic management, and land use management (Luca et al., 2020).

While most DT projects replicate physical environments and systems, there are projects that are trying to replicate human cognitive processes in urban environments (Du et al., 2020). Such projects are attempting to invent methodologies that would allow modelers to create realistic simulations of human behaviors—create DTs of humans. However, capturing human behavior via IoT sensor deployed in the city can be a challenging process, from both ethical and technical points of view, because the behavior of a human DT needs to be based on user feedback and recorded patterns, not simply on measured data (Graessler and Poehler, 2017). In this light, the Covid-19 pandemic experience opens a forum for a heated discussion, because fine-grained data recorded by the contact-tracing applications provides a perfect repository of behavioral data based on which human behavior can be modeled, yet the ethicality and the ownership right domains of this issue make the usage of these data complicated.

The notion of human DTs is mainly discussed in the healthcare literature (Liu et al., 2019), where human DTs are understood as "representations of an individual that dynamically reflect molecular status, psychological status, and lifestyle over time" (Bruynseels et al., 2018). Some pilot projects have focused on creating human DTs with the usage of wearable fitness bracelets SmartFit, through which the behavioral data about activities, food consumption, and mood has been aggregated (Barricelli et al., 2020). Other work has focused on employing human DTs in industrial settings (Amenyo, 2018; Sparrow et al., 2019).

Another proposal for the human DT framework is focused on human–computer interaction (Hafez, 2020), where the human DT is understood as a meta-model that navigates the behavioral patterns of human interaction with numerous smart machines. The focus on mimicking behavioral patterns of humans can become another milestone in human DT creation as it can potentially help to not only model anatomical and physiological processes but also enrich such models with the addition of cognitive simulations (Kawamura, 2019).

Urban mobility simulation is a part of the current generation of CDTs (Dembski et al., 2020), where agent-based models are being used to simulate the mobility behavior of city inhabitants based on mobile phone data (Wu et al., 2019). As mentioned before, the human mobility data come via three main channels: mobile phone or vehicle GPS data, cellular network connection data, or posts from social media with geo-tags (Luca et al., 2020). Agent-based modeling of human mobility in the city requires fine-grained micro-level data for input; however, this type of data is often not available due to numerous reasons including privacy concerns (Bektas and Schumann, 2019). Another drawback of the agent-based modeling approaches is that they are not properly predictive (GeoTwin, 2020) and unable to incorporate real-time data flows for short-term predictions (Kieu et al., 2020).

Another way to gather the human mobility data is to deploy IoT sensors in the city, which will allow for the passive collection of massive amounts of data (Rathore et al., 2018). This type of data would not be representative of individual travel behaviors, rather it would provide a continuously updated travel behavior of the inhabitants of the whole region of a city (Lim et al., 2018). The typical way to analyze urban Big Data is to deploy machine learning, particularly deep learning, approaches (Toch et al., 2019).

Machine learning approaches, therefore, have very strong predictive capabilities as they tend to learn the general rules of the data and derive predictions from that (Ebel et al., 2020). The natural limitation of this approach, however, is that these models may not suit well for predicting the situations which have not occurred before, because of the features of the historical data that the model has been trained on (GeoTwin, 2020).

Thus, despite the abundance of urban Big Data, the predictive power of machine learning models when trained on historical data is limited by the events that have occurred before, which makes it not applicable for certain situations, when a completely new intervention or an emergency situation is being simulated. Agent-based modeling approaches, while performing better at providing plausible scenarios for situations that have not occurred before, still are far less effective at their predictive power, as well as require sensitive data for its simulations.

4. The Potential of Synthetic Data Application in CDTs

One way of approaching this challenge can be found in generating synthetic data. In broad understanding, synthetic data are artificial data, which have characteristics similar to real data. Mainly used in cases where the real data are sensitive, it can also be used for replacing missing data and augmentation of artificial data with real data (Kaloskampis, 2019). With vast possible applications of this technology, currently, it is being utilized in order to generate useful private data without violating the data privacy laws (Bellovin et al., 2019).

Fully synthetic data can be understood as artificial data that are statistically similar to the original data (Park et al., 2018), but the new data have no identifiable information about their origin Dankar and Ibrahim (2021). This approach breaks the links between the original data and the new synthetic data so that reidentification is not meaningful anymore (Taub et al., 2018). This approach is believed to be secure for individual privacy, as it does not map back to the information about real individuals (Hu, 2018).

While real behavior data are valuable and sensitive, in many domains, this type of data is unavailable, which makes the usage of synthetic data solutions attractive (Nikolenko, 2019). In most cases, the models trained on the synthetic data are almost as effective as models trained on the real data (Hittmeir et al., 2019). Synthetic data approaches are popular in healthcare, because of the obvious concerns for data sensitivity and privacy, where the models trained on synthetic data show only small decreases in accuracy when compared to the models trained on the real data (Rankin et al., 2020).

The general working logic of the process of synthetic data generation using available synthetic data generators is that the generator takes a dataset with private information as an input, constructs a statistical model of the statistical properties of the data, and then uses this model to generate synthetic datasets that are statistically similar to the original dataset but have no identifiable information from the original data (Dankar and Ibrahim, 2021).

Another key feature of synthetic data is that it allows the creation of data about hypothetical events that have not yet happened in real life (Nikolenko, 2019). This approach is currently used in the autonomous vehicle industry, where the developers are testing very unlikely yet possible scenarios that a vehicle can encounter in the city with the usage of synthetic data. For example, "a reflective flatbed crossing a highway at dusk, with the sun's glare rendering it unintelligible to visual sensors trained only on daylight at noon" (Atherton, 2019). Other understanding of the synthetic data usage for machine learning algorithms training in the domain of autonomous vehicles includes aggregating the data from computer games in order to combine it with the real data, which helps to save both money and time for the developers (Richter et al., 2016).

Learning to drive a vehicle is a challenging process, which requires the application of reinforcement learning (RL) techniques so that an agent can learn from interacting with the environment—making real-life experiments in this domain expensive and impractical (Nikolenko, 2019). This problem is actively tackled with the usage of synthetic data, where the models are being trained in computer-generated 3D environments instead of the real world. Importantly, the performance difference between the algorithms trained in the real world and in the virtual world are almost nonexistent (Gaidon et al., 2016).

Johnson-Roberson et al. (2017) used the data from a video game Grand Theft Auto 5 with sufficiently realistic graphics in order to train the model on synthetic data from scratch, which allowed them to make a model trained in a rich virtual world to recognize and classify real objects using synthetic data. A similar approach has been exercised by Ros et al. (2016), who recreated New York City in Unity platform and then rendered more than 200,000 synthetic images of the city, which can be used for the autonomous vehicle training. This approach has been further applied to other cities, for example, San Francisco (Hernandez-Juarez et al., 2017). Li et al. (2017) used the same approach in order to create a dataset of synthetic images of foggy conditions in the city.

Datasets with 3D synthetic data of outdoor environments are less common (Nikolenko, 2019); however, some attempts of adding a LiDAR simulation into the Grand Theft Auto V video game have been made, and synthetic data was generated via this method (Wu et al., 2018). A similar approach has been exercised by Forensic Architecture (2018), where the synthetically generated images of a tank in different environments were created in order to train a machine learning model which was supposed to look for tank appearances in videos posted on social media—a sign of a potential human rights violation.

Video games have been historically used as a prominent source of virtual environments for the development of RL and other AI techniques (Schaul et al., 2011; Justesen et al., 2019). Games like Doom (Bhatti et al., 2016) and Minecraft (Oh et al., 2016) have been used to train the algorithms for robotic navigation in complex synthetic environments (Nikolenko, 2019). Real-time strategy games, such as StarCraft, have also been previously used for the generation of synthetic datasets (Lin et al., 2017), while autonomous driving agents have been trained in different racing game simulators (Sulkowski et al., 2018).

However, the agents that are being trained in the virtual environments of video games usually are not transferred directly to the real world. "There is usually no goal to transfer, say, an RL (reinforcement learning) agent playing StarCraft to a real armed conflict (thankfully)" (Nikolenko, 2019), although with the advancement of technology such transferring will, most likely, become easier.

Synthetic data generation approaches have also been used in the context of urban management, where synthetic trajectories with realistic mobility patterns have been generated (Feng et al., 2020). Synthetic trajectories are crucial for urban planning (Kang et al., 2021), computational epidemiology (Cárcamo et al., 2017), and other types of tasks that require "what if" simulation, for example, changes in mobility patterns in the presence of new infrastructure or terrorist attacks (Luca et al., 2020). Synthetic data approaches to urban mobility data also show their effectiveness in protecting the privacy of trajectory micro-data (Fiore et al., 2020).

Agent-based modeling techniques are often applied to the modeling of the social layer of the urban models, as these approaches allow for the behavioral simulation on an individual level, where the behavior of an agent is determined by its attributes on which its predictive capabilities are dependent (Chapuis and Taillandier, 2019). These agents form a synthetic population of the model—a simplified microscopic representation of a real population (Antoni et al., 2017), which only keeps the attributes that are of interest to the model (Ziemke et al., 2016). Similar to the broader discussion of the synthetic data, synthetic population matches the aggregated statistical measures of the real population, without reproducing every single unit of the population (Lenormand and Deffuant, 2013).

The synthetic population approaches have shown to be effective in epidemic simulation applications (Wu et al., 2018) and in deriving synthetic population from the country-wide census for privacy protection of the individuals (Wickramasinghe et al., 2020). This approach also has vast policy implications, because by using synthetic populations, policymakers can evaluate the city-scale built environment policies (He et al., 2020), analyze the risk for cardiovascular disease (Krauland et al., 2020), and evaluate electricity consumption in a neighborhood (van Dam et al., 2017).

This wide spectrum of applications makes the approach of synthetic population derivation from the micro-level urban mobility data a potentially effective addition to the CDTs, as it will allow for both the simulation with individual-level data without violating privacy laws, as well as provide an instrument for "what-if" simulations about the events for which there are currently no available data. This fact can be especially relevant for policymakers working with CDTs, as data may simply not yet exist for some of the

policy proposals that they want to test in a virtual urban environment (AImultiple, 2018). Creating artificial behavioral models of human behaviors in the city can be an important next step in the utilization of this technology, allowing policymakers to avoid falling into the trap of assuming that people will react to new initiatives similarly to how they reacted in the past (Levina and Duerk, 2018).

However, the effectiveness of the model with synthetic data is very dependent on the level of accuracy to which it represents the attributes of the real data (Chapuis and Taillandier, 2019). As mentioned before, for certain potential policy interventions which may be of interest to policymakers working with CDTs, the data about the behavioral response to these interventions may not exist. Thus, purely synthesized data may not satisfy the development of the model for predicting the outcome of a certain intervention.

5. A Task-Based Approach to Urban Mobility Data Generation

In order to tackle this issue, a new approach for data generation is proposed. This approach can be utilized in the Virtual Singapore project discussed in detail earlier in this article. This proposal is based on the idea that the understanding of synthetic behavioral data should be expanded from the artificial data generated by the computer which retains the values of the real data, to the artificial data generated by humans which mimics the real data as if it was generated. This approach can be applied for situations when the data are needed about the potential behavioral response to the activities that have not yet happened and for which the data are not available but needed.

The crucial question in this proposal, however, is what methodologies can be applied for the generation of such fine-grained behavioral data, based on which synthetic models of human behaviors could be made, combined, and turned into synthetic populations of CDTs.

The approach that we propose in this article is informed by the data-labeling practices (Murgia, 2019) currently exercised by technological corporations. Data labeling service providers ask employees to conduct simple tasks, which cannot be executed by computers for the creation of data on which machine learning models will be trained. The services include a comparison of two images, search results, and interface designs; representativeness of search query with the search results; search of information on the Internet; and field tasks, such as checking if the business is still in operation or secretly buying a product from a store and writing a review. These services constitute the backbone of the AI revolution (GAHNTZ, 2018), as they provide services crucial to the training of any AI system and employ millions (Reese and Heath, 2016) of blue-collar workers (Reese, 2016) around the planetary-scale labor market (Graham, 2018).

Similar methodologies are discussed in the literature about games with a purpose, where manipulations with the real data are an integral part of the gameplay, as well as in urban data games, where the game content is based on real-world data. Data games refer to the games in which the game content is based on real data and the gameplay revolves around exploration and learning from the data (Friberger et al., 2013). The design of CDTs provides a perfect ground for the design of data games, as the data used for the creation of the model of the city is the real-world data that was compiled into a model. Urban data games have been previously used for increasing the data literacy among urban populations (Wolff et al., 2015; 2017), which can be relevant in the context of CDTs, where urban data games can be created in order to involve the citizens in the process of understanding the data layers the city consists of.

The methodology of games with a purpose, games in which people create data or perform manipulation over data which computers cannot do as a side effect of the play (von Ahn and Dabbish, 2008), can become another reference model for this proposal. In this sense, the act of playing should be treated as an act of data generation. These ideas can be extrapolated and taken outside of web-service optimization to the domain of public policy data generation.

Taking these ideas further, a Task-Based Approach to Urban Mobility Data Generation is proposed. Under this approach, city inhabitants would be asked to conduct certain activities in the city in order to generate data about the activities that have not yet happened but could happen. While the potential application of this methodology can be broad, urban mobility data generation seems to be the domain in which this approach can be applied first. Such services can ask people to conduct certain activities in the city—for example, going from place A to place B by using only certain types of transportation, asking

people to imagine how they might have behaved under emergency situations with a set of parameters, or generating behavioral data in locations in which other sources of such data (GPS, cellular, or geo-tagged social media posts) are scarce. This approach can potentially become a valuable source of data for a situation in which reliance on historical data does not provide accurate predictions about certain hypothetical events, as well as for a situation in which micro-level individual data are needed for agent-based modeling purposes, but the data cannot be utilized due to privacy complications. This approach to urban mobility data generation does not violate any privacy legislation, because the data that are being generated by a person are about a hypothetical behavioral reaction to an intervention, but since the data are generated via specific tasks, they do not represent the behavior of a real individual, as the recorded behavior did not occur naturally.

As currently synthetic data are mainly used in cases where the real data are very sensitive and the methodology of synthetic data allows for anonymization of data through derivation of the data points that resemble the real data, the simulation games can provide an opportunity for anonymization on two levels. First, because the scenarios are hypothetical, the responses are not very reflective of any single individual. Second, the process of normal anonymization through synthetic data creation can be applied to this data in order to create second-order synthetic behavioral data, where the data values aggregated through simulation games can be approximated again.

Another potential benefit of this approach is that it allows the creation of a medium through which city inhabitants can be engaged in the process of urban policymaking, as they will be providing the data responses to certain proposals for intervention. This data response can serve as a medium through which the citizens can engage in the discussion about urban policy and design and express their opinion through their data responses. The approach can potentially be valuable for both policymakers and city residents, where the former will be provided with the data responses for the potential interventions for which the historical data does not suffice, while the latter will have a channel through which they can express their attitude toward the proposals.

Citizen upskilling and labor provision can be another way in which this approach can contribute to city development because, following the model of data-labeling services, city-level services for task-based data generation can be established. These services can become a vehicle for delivering public value through more-informed policies and citizen engagement.

In the case of Virtual Singapore, currently, the predictive capability of this CDT is limited by overreliance on historical data. As has been previously discussed, there are types of problems for which neither agent-based simulations nor machine learning approaches are helpful either because of the lack of data or because of privacy concerns associated with this data. Modeling of realistic human mobility and urban behavior in the times of the Covid-19 pandemic is one of these issues.

The utilization of the approach introduced in this article within the simulative environment of the CDTs, like Virtual Singapore, can allow for the creation of synthetic datasets representative of human behavioral response to hypothetical conditions. In terms of urban mobility data, under the conditions of the pandemic, this approach can be used for the approximation of the preferable mode of transportation in unusual situations for which the historical data is unavailable.

As argued in the McKinsey report (Hattrup-Silberberg et al., 2021), reducing the risk of infection has become the most important reason for choosing the mode of transportation in the city during the pandemic (time to the destination was the prime reason before the pandemic). Different cities had different experiences with the way the pandemic affected urban mobility patterns (Gragera et al., 2021). For example, Singapore has experienced a cycling boom (Abdullah, 2020) and a subsequent spike in cycling accidents (Abdullah, 2021).

While estimating the true effects of extraordinary events like the Covid-19 pandemic and designing a response strategy to it is very hard, a task-based approach introduced in this article can be helpful for gathering behavioral data about unlikely yet possible scenarios. In the case of urban mobility during the pandemic, this approach could have been utilized in a way that the respondents were asked about the way they would get from point A to point B in the scenario of social distancing. This behavioral response could have provided the data about population's preferences in terms of the mode of transportation by showing if

people would become more reliant than usual on driving personal cars, cycling, or using electric scooters (in circumstances where previously they would have taken a public transport), as well as show people's attitudes toward using car-sharing and bike-sharing services in a time when personal hygiene is of priority importance.

The insights generated from analyzing these behavioral responses can be relevant when designing a governmental strategy for urban mobility during the pandemic, as it will show to which areas the resources should be allocated and certain risks (such as with a spike of bicycle accidents) can be expected and potentially mitigated. While it is obviously impossible to gather data responses for all possible situations, the application of this approach is still much more realistic in terms of the resources needed than running a full-scale policy experiment on an urban scale.

There are several potential challenges associated with the utilization of this approach, which will need to be addressed. First, an incentive scheme needs to be established, at least at the early stages of the utilization of this approach, in order to encourage people's participation. Second, it is highly likely that some groups of populations will be more interested and active in participating in this approach than other groups. Third, some participants can have biased views on certain situations, which can affect the quality of data. Fourth, it is possible that certain tasks will be attracting more participants than others, leading to the reduced quantity of gathered information. Finally, there are potential challenges in reaching out to and including minor communities for guaranteeing the inclusion of their opinions through this approach.

6. Conclusion

This article investigates the potential that the technology of CDTs has for simulating policy interventions. As policy responses to contemporary challenges require both planning and fast reaction, new tools for envisioning the future are needed. CDTs have the characteristics which can provide adequate platforms for such scenario planning.

However, as can be seen from the example of the Virtual Singapore project, current CDTs are mainly made for the manipulations with the built environment, thus providing a toolset predominantly aimed at the physicality of the city. While the simulations of the activities conducted by city populations in urban environments are parts of the CDT models, they are usually based on historical data, which limits the predictive capabilities for a unique situation for which data are unavailable, as well as poses certain threats to the privacy of the data collected about individuals. Due to these reasons, the usage of fine-grained micro-level behavioral data for governmental simulations is complicated. Synthetic data are believed to be a potential answer to these challenges.

In the context of this article, the idea of synthetic data is expanded from artificial data generated by the computer which retains the values of the real data to the artificial data generated by humans which mimic the real data as if they were generated. Based on this notion, a Task-Based Approach to Urban Mobility Data Generation is proposed. This approach is informed by the practices of data labeling, urban data games, and games with a purpose. Under this approach, the government may ask city inhabitants to conduct certain tasks in the city in order to generate fictional data, which would provide some insights about how people would react to a hypothetical policy intervention for which there exists no data yet as such intervention has never happened before. Based on these behavioral patterns, the government can create a synthetic population for the CDT which would provide information about a realistic behavioral response to a hypothetical policy intervention without violating any legal or privacy concerns, as the data generated through this method will not represent any real individual in society.

Abbreviations

AI   artificial intelligence
NASA   The National Aeronautics and Space Administration

References

Abdullah Z (2020) Singapore sees cycling boom amid COVID-19, with increased ridership and bicycle sales. CNA. Available at https://www.channelnewsasia.com/news/singapore/covid-19-cycling-popularity-bicycle-sales-shared-bikes-13034350 (accessed 14 June 2021).

Abdullah Z (2021) Spike in bicycle accidents on Singapore roads amid cycling boom last year. CNA. Available at https://www.channelnewsasia.com/news/singapore/bicycle-road-accidents-spike-amid-cycling-boom-14986064 (accessed 14 June 2021).

AImultiple (2018) The ultimate guide to synthetic data in 2020, July 19. Available at https://research.aimultiple.com/synthetic-data/ (accessed 11 December 2020).

Amenyo J-T (2018) Using digital twins and intelligent cognitive agencies to build platforms for automated CxO future of work. arXiv:1808.07627 [Cs]. Available at http://arxiv.org/abs/1808.07627 (accessed 10 December 2020).

Antoni J-P, Vuidel G and Klein O (2017) *Generating a Located Synthetic Population of Individuals, Households, and Dwellings* (SSRN Scholarly Paper ID 2972615). Luxembourg: Social Science Research Network. https://doi.org/10.2139/ssrn.2972615.

Atherton KD (2019) Is synthetic data the key to unlocking automated war? December 27. Available at https://www.c4isrnet.com/artificial-intelligence/2019/12/27/is-synthetic-data-the-key-to-unlocking-automated-war/ (accessed 11 December 2020).

Barbosa-Filho H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, Menezes R, Ramasco JJ, Simini F and Tomasini M (2017) Human mobility: Models and applications. arXiv:1710.00004 [Physics]. https://doi.org/10.1016/j.physrep.2018.01.001.

Barricelli BR, Casiraghi E, Gliozzo J, Petrini A and Valtolina S (2020) Human digital twin for fitness management. *IEEE Access 8*, 26637–26664. https://doi.org/10.1109/ACCESS.2020.2971576.

Batty M (2018) Digital twins. *Environment and Planning B: Urban Analytics and City Science 45*(5),817–820. https://doi.org/10.1177/2399808318796416.

Bektas A and Schumann R (2019) Using mobility profiles for synthetic population generation, September 22.

Bellovin SM, Dutta PK and Reitinger N (2019) Privacy and synthetic datasets. *Stanford Technology Law Review 22*,1.

Bhatti S, Desmaison A, Miksik O, Nardelli N, Siddharth N and Torr PHS (2016) Playing doom with SLAM-augmented deep reinforcement learning. arXiv:1612.00380 [Cs, Stat]. Available at http://arxiv.org/abs/1612.00380 (accessed 21 April 2021).

Biesinger F and Weyrich M (2019) The facets of digital twins in production and the automotive industry. In *2019 23rd International Conference on Mechatronics Technology (ICMT)*. Salerno: IEEE, pp. 1–6. https://doi.org/10.1109/ICMECT.2019.8932101.

Blanford JI, Huang Z, Savelyev A and MacEachren AM (2015) Geo-located tweets. Enhancing mobility maps and capturing cross-border movement. *PLoS One 10*(6), e0129202. https://doi.org/10.1371/journal.pone.0129202.

Bliss L (2019) Why real-time traffic control has mobility experts spooked. Bloomberg.Com, July 19. Available at https://www.bloomberg.com/news/articles/2019-07-19/why-cities-want-digital-twins-to-manage-traffic (accessed 20 April 2021).

Boje C, Guerriero A, Kubicki S and Rezgui Y (2020) Towards a semantic construction digital twin: Directions for future research. *Automation in Construction 114*, 103179. https://doi.org/10.1016/j.autcon.2020.103179.

Boschert S and Rosen R (2016) Digital twin—The simulation aspect. In Hehenberger P and Bradley D (eds), *Mechatronic Futures: Challenges and Solutions for Mechatronic Systems and their Designers*. Cham: Springer International Publishing, pp. 59–74. https://doi.org/10.1007/978-3-319-32156-1_5.

Bruynseels K, Santoni de Sio F and van den Hoven J (2018) Digital twins in health care: Ethical implications of an emerging engineering paradigm. *Frontiers in Genetics 9,* 31. https://doi.org/10.3389/fgene.2018.00031.

Burbey I and Martin T (2012) A survey on predicting personal mobility. *International Journal of Pervasive Computing and Communications 8,* 5–22. https://doi.org/10.1108/17427371211221063.

Cárcamo JG, Vogel RG, Terwilliger AM, Leidig JP and Wolffe G (2017) Generative models for synthetic populations. In *Proceedings of the Summer Simulation Multi-Conference*. Ontario: Association for Computing Machinery, pp. 1–12.

Chapuis K and Taillandier P (2019) A brief review of synthetic population generation practices in agent-based social simulation, September 27.

Chen X, Kang E, Shiraishi S, Preciado VM and Jiang Z (2018) Digital behavioral twins for safe connected cars. In *Proceedings of the 21th ACM/IEEE International Conference on Model Driven Engineering Languages and Systems*. New York: ACM, pp. 144–153. https://doi.org/10.1145/3239372.3239401.

Cioara T, **Anghel I**, **Antal M**, **Salomie I**, **Antal C and Ioan AG** (2021) An overview of digital twins application domains in smart energy grid. arXiv: 2104.07904 [Cs, Eess]. Available at http://arxiv.org/abs/2104.07904 (accessed 21 April 2021).

Dankar FK and Ibrahim M (2021) Fake it till you make it: Guidelines for effective synthetic data generation. *Applied Sciences 11* (5), 2158. https://doi.org/10.3390/app11052158.

Dassault Systemes (2018) Meet Virtual Singapore, the city's 3D digital twin. GovInsider, January 29. Available at https://govinsider.asia/digital-gov/meet-virtual-singapore-citys-3d-digital-twin/ (accessed 11 December 2020).

Dembski F, **Wössner U**, **Letzgus M**, **Ruddat M and Yamu C** (2020) Urban digital twins for smart cities and citizens: The case study of Herrenberg, Germany. *Sustainability 12*(6), 2307. https://doi.org/10.3390/su12062307.

Dembski F, **Wössner U and Yamu C** (2019) Digital twin, virtual reality and space syntax: Civic engagement and decision support for smart sustainable cities, July 12.

Deng T, **Zhang K and Shen Z-J** (2021) A systematic review of a digital twin city: A new pattern of urban governance toward smart cities. *Journal of Management Science and Engineering*. https://doi.org/10.1016/j.jmse.2021.03.003.

Deren L, **Wenbo Y and Zhenfeng S** (2021) Smart city based on digital twins. *Computational Urban Science 1*(1), 4. https://doi.org/10.1007/s43762-021-00005-y.

Dharmani S and Lulla S (2020) How digital twins give automotive companies a real-world advantage. Available at https://www.ey.com/en_us/advanced-manufacturing/how-digital-twins-give-automotive-companies-a-real-world-advantage (accessed 11 December 2020).

Dogan Ö, **Sahin O and Karaarslan E** (2021) Digital twin based disaster management system proposal: DT-DMS. arxiv: 2103.17245 [Cs]. Available at http://arxiv.org/abs/2103.17245 (accessed 21 April 2021).

Dou SQ, **Zhang HH**, **Zhao YQ**, **Wang AM**, **Xiong YT and Zuo JM** (2020) Research on construction of spatio-temporal data visualization platform for GIS and BIM fusion. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-3-W10*, 555–563. https://doi.org/10.5194/isprs-archives-XLII-3-W10-555-2020.

Du J, **Zhu Q**, **Shi Y**, **Wang Q**, **Lin Y and Zhao D** (2020) Cognition digital twins for personalized information systems of smart cities: Proof of concept. *Journal of Management in Engineering 36*(2), 04019052. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000740.

Ebel P, **Göl IE**, **Lingenfelder C and Vogelsang A** (2020) Destination prediction based on partial trajectory data. arxiv:2004.07473 [Cs, Stat]. Available at http://arxiv.org/abs/2004.07473 (accessed 22 April 2021).

Ebrahimpour Z, **Wan W**, **Cervantes O**, **Luo T and Ullah H** (2019) Comparison of main approaches for extracting behavior features from crowd flow analysis. *ISPRS International Journal of Geo-Information 8*(10), 440. https://doi.org/10.3390/ijgi8100440.

Enders M and Hoßbach N (2019). Dimensions of digital twin applications—A literature review. *AMCIS 2019 Proceedings*. Available at https://aisel.aisnet.org/amcis2019/org_transformation_is/org_transformation_is/20 (accessed 11 December 2020).

Fan C, **Zhang C**, **Yahja A and Mostafavi A** (2021) Disaster city digital twin: A vision for integrating artificial and human intelligence for disaster management. *International Journal of Information Management 56*, 102049. https://doi.org/10.1016/j.ijinfomgt.2019.102049.

Feng J, **Yang Z**, **Xu F**, **Yu H**, **Wang M and Li Y** (2020) Learning to simulate human mobility. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York: ACM, pp. 3426–3433. https://doi.org/10.1145/3394486.3412862.

Ferguson S (2020) Apollo 13: The first digital twin, April 14. Available at https://blogs.sw.siemens.com/simcenter/apollo-13-the-first-digital-twin/ (accessed 11 December 2020).

Fiore M, **Katsikouli P**, **Zavou E**, **Cunche M**, **Fessant F**, **Hello DL**, **Aivodji UM**, **Olivier B**, **Quertier T and Stanica R** (2020) Privacy in trajectory micro-data publishing: A survey. arxiv:1903.12211 [Cs]. Available at http://arxiv.org/abs/1903.12211 (accessed 21 April 2021).

Forensic Architecture (2018) Synthetic data generation: Development of data calssification tools. Forensic architecture. Available at https://synthetic-datasets.sfo2.digitaloceanspaces.com/website/FA_Synthetic-Data-Generation_Report.pdf (accessed 10 December 2020).

Francisco A, **Mohammadi N and Taylor JE** (2020) Smart city digital twin–enabled energy management: Toward real-time urban building energy benchmarking. *Journal of Management in Engineering 36*(2), 04019045. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000741.

Friberger M, **Togelius J**, **Borg Cardona A**, **Ermacora M**, **Mousten A**, **Møller Jensen M**, **Tanase V-A and Brøndsted U** (2013) Data games, pp. 1–8. Available at http://urn.kb.se/resolve?urn=urn:nbn:se:mau:diva-12543 (accessed 10 December 2020).

Fuller A, **Fan Z**, **Day C and Barlow C** (2020) Digital twin: Enabling technologies, challenges and open research. *IEEE Access 8*, 108952–108971. https://doi.org/10.1109/ACCESS.2020.2998358.

Gahntz M (2018) The invisible workers of the AI era. Medium, December 20. Available at https://towardsdatascience.com/the-invisible-workers-of-the-ai-era-c83735481ba (accessed 10 December 2020).

Gaidon A, **Wang Q**, **Cabon Y and Vig E** (2016) Virtual worlds as proxy for multi-object tracking analysis. arxiv:1605.06457 [Cs, Stat]. Available at http://arxiv.org/abs/1605.06457 (accessed 21 April 2021).

Geddie J and Aravindan A (2018) Virtual Singapore project could be test bed for planners—And plotters. Reuters, September 28. Available at https://www.reuters.com/article/us-singapore-technology-idUSKCN1M70U1 (accessed 12 December 2020).

GeoTwin (2020) Agent-based models as digital twins, the present and the future. Available at https://geotwin.io/en/company/newsroom/blog/agent-based-models-as-digital-twins-the-present-and-the-future (accessed 10 December 2020).

Glaessgen E and Stargel D (2012) The digital twin paradigm for future NASA and U.S. air force vehicles. In *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*. Reston, VA: American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2012-1818.

Gobeawan L, **Lin E**, **Tandon A**, **Yee A**, **Khoo V**, **Teo S**, **Yi S**, **Lim C**, **Wong S**, **Wise D**, **Cheng P**, **Liew SC**, **Huang X**, **Li Q**, **Teo L**, **Fekete GS and Poto M** (2018) Modeling trees for virtual Singapore: From data acquisition to citygml models. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-4/W10*, 55–62. https://doi.org/10.5194/isprs-archives-XLII-4-W10-55-2018.

González MC, **Hidalgo CA and Barabási A-L** (2008) Understanding individual human mobility patterns. *Nature 453*(7196), 779–782. https://doi.org/10.1038/nature06958.

Graessler I and Poehler A (2017) Integration of a digital twin as human representation in a scheduling procedure of a cyber-physical production system. In *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. Singapore: IEEE, pp. 289–293. https://doi.org/10.1109/IEEM.2017.8289898.

Gragera A, **Albalate D**, **Bel G**, **Schaj G**, **Cañas H**, **Aquilué I**, **Helder J**, **Espindola L**, **Mósca M**, **Edelstam M**, **Marti M**, **Shetty N**, **Barton P**, **Riegebauer P**, **Filohn P and Urbano R** (2021) Urban mobility strategies during COVID-19. Available at https://www.eiturbanmobility.eu/wp-content/uploads/2021/03/Urban-mobility-strategies-during-COVID-19_long-2.pdf (accessed 14 June 2021).

Graham M (2018) The rise of the planetary labour market – And what it means for the future of work. NS Tech. Available at https://tech.newstatesman.com/guest-opinion/planetary-labour-market (accessed 10 December 2020).

Grieves M and Vickers J (2017) Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In Kahlen F-J, Flumerfelt S and Alves A (eds), *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches*. New York: Springer International, pp. 85–113. https://doi.org/10.1007/978-3-319-38756-7_4.

Guerrini F (2016) 'Virtual Singapore' platform will help the city-state address shortage of space, aging population. Forbes. Available at https://www.forbes.com/sites/federicoguerrini/2016/08/18/virtual-singapore-platform-will-help-the-city-state-address-shortage-of-space-aging-population/ (accessed 10 December 2020).

Hafez W (2020) Human digital twin: Enabling human-multi smart machines collaboration. In Bi Y, Bhatia R and Kapoor S (eds), *Intelligent Systems and Applications*. Cham: Springer International, pp. 981–993. https://doi.org/10.1007/978-3-030-29513-4_72.

Ham Y and Kim J (2020) Participatory sensing and digital twin city: Updating virtual city models for enhanced risk-informed decision-making. *Journal of Management in Engineering 36*(3), 04020005. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000748.

Hattrup-Silberberg M, **Hausler S**, **Heineke K**, **Laverty N**, **Möller T**, **Schwedhelm D and Wu T** (2021) Five COVID-19 aftershocks reshaping mobility's future | McKinsey. Available at https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/five-covid-19-aftershocks-reshaping-mobilitys-future# (accessed 14 June 2021).

He BY, **Zhou J**, **Ma Z**, **Chow JYJ and Ozbay K** (2020) Evaluation of city-scale built environment policies in New York City with an emerging-mobility-accessible synthetic population. *Transportation Research Part A: Policy and Practice 141*, 444–467. https://doi.org/10.1016/j.tra.2020.10.006.

Hemetsberger L (2020) Cities & digital twins: From hype to reality, May 26. Available at https://oascities.org/three-key-challenges-towards-digital-twin-adoption-at-scale/ (accessed 11 December 2020).

Hernandez-Juarez D, **Schneider L**, **Espinosa A**, **Vázquez D**, **López AM**, **Franke U**, **Pollefeys M and Moure JC** (2017) Slanted stixels: Representing San Francisco's steepest streets. arxiv:1707.05397 [Cs]. Available at http://arxiv.org/abs/1707.05397 (accessed 21 April 2021).

Hittmeir M, **Ekelhart A and Mayer R** (2019) On the utility of synthetic data: An empirical evaluation on machine learning tasks. In *Proceedings of the 14th International Conference on Availability, Reliability and Security*. New York: ACM, pp. 1–6. https://doi.org/10.1145/3339252.3339281.

Holstein W (2015) Virtual Singapore. Compass Magazine, October 27. Available at https://compassmag.3ds.com/virtual-singapore/ (accessed 11 December 2020).

Hu J (2018) Bayesian estimation of attribute and identification disclosure risks in synthetic data. arxiv:1804.02784 [Stat]. Available at http://arxiv.org/abs/1804.02784 (accessed 20 April 2021).

Ismagilova E, **Hughes L**, **Rana NP and Dwivedi YK** (2020) Security, privacy and risks within smart cities: Literature review and development of a smart city interaction framework. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-020-10044-1.

Jiang S, **Fiore GA**, **Yang Y**, **Ferreira J**, **Frazzoli E and González MC** (2013) A review of urban computing for mobile phone traces: Current methods, challenges and opportunities. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*. New York: ACM, pp. 1–9. https://doi.org/10.1145/2505821.2505828.

Jo S, **Park D**, **Park H and Kim S** (2018) Smart livestock farms using digital twin: Feasibility study. In *2018 International Conference on Information and Communication Technology Convergence (ICTC)*. Jeju: IEEE, pp1461–1463. https://doi.org/10.1109/ICTC.2018.8539516.

Johnson-Roberson M, **Barto C**, **Mehta R**, **Sridhar SN**, **Rosaen K and Vasudevan R** (2017) Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks? arxiv:1610.01983 [Cs]. Available at http://arxiv.org/abs/1610.01983 (accessed 21 April 2021).

Jones D, **Snider C**, **Nassehi A**, **Yon J and Hicks B** (2020) Characterising the digital twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology 29*, 36–52. https://doi.org/10.1016/j.cirpj.2020.02.002.

Justesen N, **Bontrager P**, **Togelius J and Risi S** (2019) Deep learning for video game playing. arxiv:1708.07902 [Cs]. Available at http://arxiv.org/abs/1708.07902 (accessed 21 April 2021).

Kaloskampis I (2019) Synthetic data for public good | Data Science Campus. Available at https://datasciencecampus.ons.gov.uk/projects/synthetic-data-for-public-good/ (accessed 23 December 2020).

Kang X, **Liu L**, **Zhao D and Ma H** (2021) TraG: A trajectory generation technique for simulating urban crowd mobility. *IEEE Transactions on Industrial Informatics 17*(2), 820–829. https://doi.org/10.1109/TII.2020.2976777.

Kawamura R (2019) A digital world of humans and society—Digital twin computing. Available at https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202003fa1.pdf&mode=show_pdf (accessed 11 December 2020).

Kieu L-M, **Malleson N and Heppenstall A** (2020) Dealing with uncertainty in agent-based models for short-term predictions. *Royal Society Open Science 7*(1), 191074. https://doi.org/10.1098/rsos.191074.

Klebanov B, **Antropov T and Zvereva O** (2019) Hybrid automaton implementation for intelligent agents' behavior modelling. In *2019 International Conference on Information Science and Communications Technologies (ICISCT)*. Tashkent: IEEE, pp. 1–4. https://doi.org/10.1109/ICISCT47635.2019.9011955.

Koçer KK (2020) URN:NBN Resolver für Deutschland Und Schweiz. German National Library. Available at https://nbn-resolving.org/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20200907-1548861-1-4 (accessed 21 April 2021).

Krauland MG, **Frankeny RJ**, **Lewis J**, **Brink L**, **Hulsey EG**, **Roberts MS and Hacker KA** (2020) Development of a synthetic population model for assessing excess risk for cardiovascular disease death. *JAMA Network Open 3*(9), e2015047. https://doi.org/10.1001/jamanetworkopen.2020.15047.

Kritzinger W, **Karner M**, **Traar G**, **Henjes J and Sihn W** (2018) Digital twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine 51*(11), 1016–1022. https://doi.org/10.1016/j.ifacol.2018.08.474.

Kuts V, **Otto T**, **Tähemaa T and Bondarenko Y** (2019) Digital twin based synchronised control and simulation of the industrial robotic cell using virtual reality. *Journal of Machine Engineering 19*, 128–144. https://doi.org/10.5604/01.3001.0013.0464.

Laamarti F, **Badawi HF**, **Ding Y**, **Arafsha F**, **Hafidh B and Saddik AE** (2020) An ISO/IEEE 11073 standardized digital twin framework for health and well-being in smart cities. *IEEE Access 8*, 105950–105961. https://doi.org/10.1109/ACCESS.2020.2999871.

Lenormand M and Deffuant G (2013) Generating a synthetic population of individuals in households: Sample-free vs sample-based methods. *Journal of Artificial Societies and Social Simulation 16*(4),12.

Levina M and Duerk B (2018) Fidelity of the digital twin. Dell EMC. Available at https://education.dellemc.com/content/dam/dell-emc/documents/en-us/2018KS_Levina-Fidelity_of_the_Digial_Twin.pdf (accessed 11 December 2020).

Li K, **Li Y**, **You S and Barnes N** (2017) Photo-realistic simulation of road scene for data-driven methods in bad weather. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. Venice: IEEE, pp. 491–500. https://doi.org/10.1109/ICCVW.2017.65.

Lim C, **Kim K-J and Maglio PP** (2018) Smart cities with big data: Reference models, challenges, and considerations. *Cities 82*, 86–99. https://doi.org/10.1016/j.cities.2018.04.011.

Lin Z, **Gehring J**, **Khalidov V and Synnaeve G** (2017) STARDATA: A starCraft AI research dataset. arxiv:1708.02139 [Cs]. Available at http://arxiv.org/abs/1708.02139 (accessed 22 April 2021).

Liu Y, **Zhang L**, **Yang Y**, **Zhou L**, **Ren L**, **Wang F**, **Liu R**, **Pang Z and Deen MJ** (2019) A novel cloud-based framework for the elderly healthcare services using digital twin. *IEEE Access 7*, 49088–49101. https://doi.org/10.1109/ACCESS.2019.2909828.

Lu Q, **Parlikad AK**, **Woodall P**, **Don Ranasinghe G**, **Xie X**, **Liang Z**, **Konstantinou E**, **Heaton J and Schooling J** (2020) Developing a digital twin at building and city levels: Case study of West Cambridge campus. *Journal of Management in Engineering 36*(3), 05020004. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000763.

Luca M, **Barlacchi G**, **Lepri B and Pappalardo L** (2020) Deep learning for human mobility: A survey on data and models. arxiv:2012.02825 [Cs]. Available at http://arxiv.org/abs/2012.02825 (accessed 21 April 2021).

Malik AA (2021) Framework to model virtual factories: A digital twin view. arxiv:2104.03034 [Cs, Eess]. Available at http://arxiv.org/abs/2104.03034 (accessed 21 April 2021).

Medvedev A, **Fedchenkov P**, **Zaslavsky A**, **Anagnostopoulos T and Khoruzhnikov S** (2015) Waste management as an IoT-enabled service in smart cities. In Balandin S, Andreev S and Koucheryavy Y (eds), *Internet of Things, Smart Spaces, and Next Generation Networks and Systems.* Cham: Springer International, pp. 104–115. https://doi.org/10.1007/978-3-319-23126-6_10.

Menouar H, **Guvenc I**, **Akkaya K**, **Uluagac AS**, **Kadri A and Tuncer A** (2017) UAV-enabled intelligent transportation systems for the smart city: Applications and challenges. *IEEE Communications Magazine 55*(3), 22–28. https://doi.org/10.1109/MCOM.2017.1600238CM.

Mohammadi N and Taylor J (2020) Knowledge discovery in smart city digital twins, January 1. https://doi.org/10.24251/HICSS.2020.204.

Mohammadi N and Taylor JE (2017) Smart city digital twins. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*. Honolulu, HI: IEEE, pp. 1–5. https://doi.org/10.1109/SSCI.2017.8285439.

Mourtzis D (2020) Simulation in the design and operation of manufacturing systems: State of the art and new trends. *International Journal of Production Research 58*(7), 1927–1949. https://doi.org/10.1080/00207543.2019.1636321.

Murgia M (2019) AI's new workforce: The data-labelling industry spreads globally, July 24. Available at https://www.ft.com/content/56dde36c-aa40-11e9-984c-fac8325aaa04 (accessed 12 December 2020).

National Research Foundation Singapore (2018) Virtual Singapore. Available at https://www.nrf.gov.sg/programmes/virtual-singapore (accessed 23 December 2020).

Nativi S, **Delipetrev B and Craglia M** (2020) Destination Earth: Survey on "Digital Twins" technologies and activities, in the Green Deal area [Text]. EU Science Hub - European Commission, November 10. Available at https://ec.europa.eu/jrc/en/publication/destination-earth-survey-digital-twins-technologies-and-activities-green-deal-area (accessed 22 April 2021).

Nikolenko SI (2019) Synthetic data for deep learning. arxiv:1909.11512 [Cs]. Available at http://arxiv.org/abs/1909.11512 (accessed 21 April 2021).

Nochta T, **Wan L**, **Schooling JM and Parlikad AK** (2021) A socio-technical perspective on urban analytics: The case of city-scale digital twins. *Journal of Urban Technology 28*(1–2), 263–287. https://doi.org/10.1080/10630732.2020.1798177.

Oh J, **Chockalingam V**, **Singh S and Lee H** (2016) Control of memory, active perception, and action in minecraft. arxiv: 1605.09128 [Cs]. Available at http://arxiv.org/abs/1605.09128 (accessed 21 April 2021).

Oldenbroek V, **Verhoef LA and van Wijk AJM** (2017) Fuel cell electric vehicle as a power plant: Fully renewable integrated transport and energy system design and analysis for smart city areas. *International Journal of Hydrogen Energy 42*(12), 8166–8196. https://doi.org/10.1016/j.ijhydene.2017.01.155.

Pang J, **Li J**, **Xie Z Huang Y and Cai Z** (2020) Collaborative city digital twin For Covid-19 pandemic: A federated learning solution. arxiv:2011.02883 [Cs]. Available at http://arxiv.org/abs/2011.02883 (accessed 22 April 2021).

Pappalardo L, **Rinzivillo S**, **Qu Z**, **Pedreschi D and Giannotti F** (2013) Understanding the patterns of car travel. *The European Physical Journal Special Topics 215*(1), 61–73. https://doi.org/10.1140/epjst/e2013-01715-5.

Park H, **Easwaran A and Andalam S** (2019) Challenges in digital twin development for cyber-physical production systems. *arxiv: 2102.03341* [Cs], 11615, pp. 28–48. https://doi.org/10.1007/978-3-030-23703-5_2.

Park N, **Mohammadi M**, **Gorde K**, **Jajodia S**, **Park H and Kim Y** (2018) Data synthesis based on generative adversarial networks. *Proceedings of the VLDB Endowment 11*(10), 1071–1083. https://doi.org/10.14778/3231751.3231757.

Parra L, **Sendra S**, **Lloret J and Bosch I** (2015) Development of a conductivity sensor for monitoring groundwater resources to optimize water management in smart city environments. *Sensors 15*(9), 20990–21015. https://doi.org/10.3390/s150920990.

Qi Q and Tao F (2018) Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison. *IEEE Access 6*, 3585–3593. https://doi.org/10.1109/ACCESS.2018.2793265.

Rankin D, **Black M**, **Bond R**, **Wallace J**, **Mulvenna M and Epelde G** (2020) Reliability of supervised machine learning using synthetic data in health care: Model to preserve privacy for data sharing. *JMIR Medical Informatics 8*(7), e18910. https://doi.org/10.2196/18910.

Rathore MM, **Paul A**, **Hong W-H**, **Seo H**, **Awan I and Saeed S** (2018) Exploiting IoT and big data analytics: Defining smart digital city using real-time urban data. *Sustainable Cities and Society 40*, 600–610. https://doi.org/10.1016/j.scs.2017.12.022.

Reese (2016) Is 'data labeling' the new blue-collar job of the AI era? TechRepublic, 10:00 AM PST. Available at https://www.techrepublic.com/article/is-data-labeling-the-new-blue-collar-job-of-the-ai-era/ (accessed 23 April 2021).

Reese H and Heath N (2016) Inside Amazon's clickworker platform: How half a million people are being paid pennies to train AI. TechRepublic, December 16. Available at https://www.techrepublic.com/article/inside-amazons-clickworker-platform-how-half-a-million-people-are-training-ai-for-pennies-per-task/ (accessed *26 April 2021*).

Richter SR, **Vineet V**, **Roth S and Koltun V** (2016) Playing for data: Ground truth from computer games. In Leibe B, Matas J, Sebe N and Welling M (eds), *Computer Vision – ECCV 2016*. Amsterdam: Springer International, pp. 102–118. https://doi.org/10.1007/978-3-319-46475-6_7.

Ríos J, **Hernandez-Matias J**, **Oliva M and Mas F** (2015) Product avatar as digital counterpart of a physical individual product: Literature review and implications in an aircraft, July 20. https://doi.org/10.3233/978-1-61499-544-9-657.

Ros G, **Sellart L**, **Materzynska J**, **Vazquez D and Lopez AM** (2016) The SYNTHIA dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* Las Vegas, NV: IEEE, pp. 3234–3243. https://doi.org/10.1109/CVPR.2016.352.

Rožanec JM, **Lu J**, **Rupnik J**, **Škrjanc M**, **Mladenić D**, **Fortuna B**, **Zheng X and Kiritsis D** (2021) Actionable cognitive twins for decision making in manufacturing. ArXiv:2103.12854 [Cs]. Available at http://arxiv.org/abs/2103.12854 (accessed 23 April 2021).

Ruohomäki T, **Airaksinen E**, **Huuska P**, **Kesäniemi O**, **Martikka M and Suomisto J** (2018) Smart city platform enabling digital twin. In *2018 International Conference on Intelligent Systems (IS)*. Funchal: IEEE, pp. 155–161. https://doi.org/10.1109/IS.2018.8710517.

Schaul T, **Togelius J and Schmidhuber J** (2011) Measuring intelligence through games. arxiv:1109.1314 [Cs]. Available at http://arxiv.org/abs/1109.1314 (accessed 12 December 2020).

Schrotter G (2020) Towards a reliable map of subsurface utilities in Singapore digital underground.

Schrotter G and Hürzeler C (2020) The digital twin of the City of Zurich for urban planning. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science 88*(1), 99–112. https://doi.org/10.1007/s41064-020-00092-2.

Shahat E, **Hyun CT and Yeom C** (2021) City digital twin potentials: A review and research agenda. *Sustainability 13*(6), 3386. https://doi.org/10.3390/su13063386.

Sharma A, **Kosasih E**, **Zhang J**, **Brintrup A and Calinescu A** (2020) Digital twins: State of the art theory and practice, challenges, and open research questions. arxiv:2011.02833 [Cs]. Available at http://arxiv.org/abs/2011.02833 (accessed 21 April 2021).

Singapore Land Authority (2014) Annex A - Potential applications of Virtual Singapore. Available at https://www.sla.gov.sg/articles/press-releases/2014/virtual-singapore-a-3d-city-model-platform-for-knowledge-sharing-and-community-collaboration/annex-a-potential-applications-of-virtual-singapore (accessed 23 April 2021).

Soe R-M (2017) Smart twin cities via urban operating system. In *Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance*. New York: ACM, pp. 391–400. https://doi.org/10.1145/3047273.3047322.

Sparrow D, **Kruger K and Basson A** (2019) Human digital twin for integrating human workers in Industry 4.0, January 20.

Sulkowski T, **Bugiel P and Izydorczyk J** (2018) In search of the ultimate autonomous driving simulator. In *2018 International Conference on Signals and Electronic Systems (ICSES)*. Kraków, IEEE, pp. 252–256. https://doi.org/10.1109/ICSES.2018.8507288.

Tao F, **Zhang H**, **Liu A and Nee AYC** (2019) Digital twin in industry: State-of-the-art. *IEEE Transactions on Industrial Informatics* 15(4), 2405–2415. https://doi.org/10.1109/TII.2018.2873186.

Tao F and Zhang M (2017) Digital twin shop-floor: A new shop-floor paradigm towards smart manufacturing. *IEEE Access 5*, 20418–20427. https://doi.org/10.1109/ACCESS.2017.2756069.

Taub J, **Elliot M**, **Pampaka M and Smith D** (2018) Differential correct attribution probability for synthetic data: An exploration. In Domingo-Ferrer J and Montes F (eds), *Privacy in Statistical Databases.* Berlin, Germany: Springer International, pp. 122–137. https://doi.org/10.1007/978-3-319-99771-1_9.

Toch E, **Lerner B**, **Ben Zion E and Ben-Gal I** (2019) Analyzing large-scale human mobility data: A survey of machine learning methods and applications. *Knowledge and Information Systems 58,* 501–523. https://doi.org/10.1007/s10115-018-1186-x.

van Dam KH, **Bustos-Turu G and Shah N** (2017) A methodology for simulating synthetic populations for the analysis of socio-technical infrastructures. In Jager W, Verbrugge R, Flache A, deRoo G, Hoogduin L and Hemelrijk C (eds), *Advances in Social Simulation 2015*. Cham: Springer International Publishing, pp. 429–434. https://doi.org/10.1007/978-3-319-47253-9_39.

von Ahn L and Dabbish L (2008) ACM: digital library: Communications of the ACM. https://dl.acm.org/doi/fullHtml/10.1145/1378704.1378719.

Wan L, **Nochta T and Schooling JM** (2019) Developing a city-level digital twin? Propositions and a case study. In *International Conference on Smart Infrastructure and Construction 2019 (ICSIC)* (*Vol. 1–0*). Cambridge: ICE Publishing, pp. 187–194). https://doi.org/10.1680/icsic.64669.187.

White G, **Zink A**, **Codecá L and Clarke S** (2021) A digital twin smart city for citizen feedback. *Cities 110*, 103064. https://doi.org/10.1016/j.cities.2020.103064.

Wickramasinghe BN, **Singh D and Padgham L** (2020) Building a large synthetic population from Australian census data. arxiv: 2008.11660 [Cs, Stat]. Available at http://arxiv.org/abs/2008.11660 (accessed 21 April 2021).

Wolff A, **Kortuem G and Cavero J** (2015) Urban data games: Creating smart citizens for smart cities. In *2015 IEEE 15th International Conference on Advanced Learning Technologies*. Hualien: IEEE, pp. 164–165. https://doi.org/10.1109/ICALT.2015.44.

Wolff A, **Valdez A-M**, **Barker M**, **Potter S**, **Gooch D**, **Giles E and Miles J** (2017) Engaging with the smart city through urban data games. In Nijholt A (ed), *Playable Cities: The City as a Digital Playground.* Singapore: Springer, pp. 47–66. https://doi.org/10.1007/978-981-10-1962-3_3.

Wong A, **Mo C**, **Shieh K and Ee Sim S** (2020) Digital twins in smart city. PWC. Available at https://www.pwccn.com/en/research-and-insights/publications/digital-twins-in-smart-city.pdf (accessed 12 December 2020).

Wu B, **Zhou X**, **Zhao S**, **Yue X and Keutzer K** (2018) SqueezeSegV2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a LiDAR point cloud. arxiv:1809.08495 [Cs]. Available at http://arxiv.org/abs/1809.08495 (accessed 21 April 2021).

Wu H, **Liu L**, **Yu Y**, **Peng Z**, **Jiao H and Niu Q** (2019) An agent-based model simulation of human mobility based on mobile phone data: How commuting relates to congestion. *ISPRS International Journal of Geo-Information 8*(7), 313. https://doi.org/10.3390/ijgi8070313.

Wu H, **Ning Y**, **Chakraborty P**, **Vreeken J**, **Tatti N and Ramakrishnan N** (2018) Generating realistic synthetic population datasets. *ACM Transactions on Knowledge Discovery from Data 12*(4), 45.1–45.22. https://doi.org/10.1145/3182383.

Wu R, **Luo G**, **Shao J**, **Tian L and Peng C** (2018) Location prediction on trajectory data: A review. *Big Data Mining and Analytics 1*, 108–127. https://doi.org/10.26599/BDMA.2018.9020010.

Yan J, **Jaw SW**, **Soon KH**, **Wieser A and Schrotter G** (2019) Towards an underground utilities 3D data model for land administration. *Remote Sensing 11*(17), 1957. https://doi.org/10.3390/rs11171957.

Yan J, **Van Son R and Soon KH** (2021) From underground utility survey to land administration: An underground utility 3D data model. *Land Use Policy 102*, 105267. https://doi.org/10.1016/j.landusepol.2020.105267.

Zheng Y, **Xie X and Ma W-Y** (2010) GeoLife: A collaborative social networking service among user, location and trajectory. *IEEE Database Engineering Bulletin 33*, 32–39.

Ziemke D, **Nagel K and Moeckel R** (2016) Towards an agent-based, integrated land-use transport modeling system. *Procedia Computer Science 83*, 958–963. https://doi.org/10.1016/j.procs.2016.04.192.