LUND UNIVERSITY
School of Economics and Management
Department of Informatics

Algorithmic Bias in City Scale Digital Twin Data Management Processes

A qualitative exploration of the pragmatic, ethical, and moral prioritization surrounding algorithmic bias in city scale digital twin projects

Master thesis 15 HEC, course INFM10 in Information Systems

Authors: Wing Yi Lam
                    Elin Wernholm

Supervisor: Paul Pierce

Grading Teachers: Miranda Kajtazi
                    Avijit Chowdhury

Algorithmic Bias in City Scale Digital Twin Data Management Processes: A qualitative exploration of the pragmatic, ethical, and moral prioritization surrounding algorithmic bias in city scale digital twin projects

AUTHORS: Wing Yi Lam and Elin Wernholm

ABSTRACT (MAX. 200 WORDS):
The application of digital twins in the city setting is emerging, the massive amounts of data and algorithms used raise the concern of algorithmic bias. This research focuses on examining how city scale digital twin practitioners prioritize the pragmatic, ethical, and moral domains in data management processes and when encountering algorithmic bias. It aims to provide insights for practitioners and urban planners to navigate these complex challenges while considering potential trade-offs between domains. A qualitative research approach was chosen to understand decisions in data management processes during the planning and design of city scale digital twins, with semi-structured interviews as the primary data collection method. This thesis found that in data management processes and when encountering algorithmic bias in city scale digital twins, the pragmatic domain was consistently prioritized, while the ethical and moral domains received less emphasis. Trade-offs between the pragmatic and moral domains were more prevalent than those between the pragmatic and ethical domains. The findings highlight the need for an ethical code of conduct and common standards in data handling to uphold moral values, and the importance of involving all stakeholders in the discourse, including citizens.

Content

Figures

Tables

1 Introduction

1.1 Background & Context

The smart city is a growing phenomenon where the ultimate goal is to increase city resilience, sustainability and quality of life through smart innovations (Pierce, Ricciardi & Zardini, 2017). The smart city concept includes most technology-based innovations that tackle planning, development, operation, or management of urban activities (Yigitcanlar, Kamruzzaman, Buys, Ioppolo, Sabatini-Marques, da Costa & Yun, 2018). Among other things it can be utilized to catalyse efficient resource allocation (Moggi & Dameri, 2021), provide security, improve operational efficiency and facilitate human activities (Mohammadi & Taylor, 2018). The idea is that vast amounts of data is being produced from the movement and behaviour of people and communities together with the infrastructure in highly connected cities (Pierce et al., 2017). This data can be utilized to maximize services (Mohammadi & Taylor, 2018) provided in the top layer of the city (Pierce et al., 2017). One example of such a service is an application or system for reallocation of leftover food (Moggi & Dameri, 2021), another is an API from Copenhagen Solutions Lab (2018), developed to find the closest parking space, to reduce time spent looking for parking in a city.

Further technology that might contribute towards smart city goals is a digital twin. Digital twins can be defined as "real-time, virtual replicas of physical and biological entities" (Tzachor, Sabri, Richards, Rajabifard & Acuto, 2022, p.822), and are originally used in the manufacturing industry as it helps to gain efficiency during operations (Ferré-Bigorra, Casals & Gangolells, 2022; Tzachor et al., 2022). The growing interest of utilizing digital twins in cities (Tzachor et al., 2022; Evangelou, Gkeli & Potsiou, 2022; Oti-Sarpong, Bastidas, Nochta, Wan, Tang & Schooling, 2022) is sparking more holistic initiatives, as illustrated by the European Commission's partnership with the Local Digital Twin initiative (European Commission, 2021; Living in EU, n.d.a). Researchers are increasingly incorporating digital twins in urban planning and the smart city setting (Evangelou et al., 2022; Lee, Braud, Hosio & Hui, 2021; Huang, Zhang & Zeng, 2022; Herath & Mittal, 2022). The aim is to construct a system connecting the physical world with a virtual world, mapping and interacting in both directions (Deren, Wenbo & Zhenfeng, 2021). In other words, the system includes a comprehensive modelling of a city's physical pace. Starting from the terrain, layering buildings, infrastructure and mobility, capturing data about all layers of the space in a data layer (White, Zink, Codecá & Clarke, 2021).

The data captured in the data layer of a city scale digital twin can be used in the virtual system for simulations and predictions (White et al., 2021; Tzachor et al., 2022) and to improve urban planning and operations (Deren et al., 2021). Researchers document digital twin use for numerous activities and purposes, including traffic management (White et al., 2021), improvement of public services (Zhang, Chen, Zhang, Cui, Han, Meng & Xu, 2022), informing planning and resource allocation (Oti-Sarpong et al., 2022), serving environmental challenges through climate forecasts, supporting mitigation of, and adaptation to such grand issues (Tzachor et al., 2022), as well as facilitating smart energy solutions (Dulaimi, Hamida, Naser & Mawed, 2022). Digital twins utilize real time data (Batty, 2018) to provide forecasts and decision-making basis (Tzachor et al., 2022). The essence of the digital twin, or virtual layer in a smart city is using data modelling and analytics to help prioritize activities in city planning (Oti-Sarpong et al., 2022).

The expanding extent to which algorithmic processes are being utilized make its biases increasingly important to address (Zou & Schiebinger, 2018). With AI, machine learning and other tools used in digital twins (Tzachor et al., 2022) comes a risk of bias (Kordzadeh & Ghasemaghaei, 2022; Tzachor et al., 2022; Zou & Schiebinger, 2018). "[A] major driver of bias in AI is the training data. Most machine-learning tasks are trained on large, annotated data sets" (Zou & Schiebinger, 2018, p.325). Algorithmic processes are trained with society's pre-existing biases and risk continuously maintaining them (Kordzadeh & Ghasemaghaei, 2022). Besides, data is commonly used for actuarial prediction as people believe that the decisions they make are calculated on the basis of algorithms that produce unbiased output and scientific truth, in contrast to clinical prediction which tends to be viewed as more subjective (Madsen, Grundtvig & Thorsen, 2022; Brauneis & Goodman, 2018). However, this "over-reliance" on data models might result in over-simplified outcomes which risk "reinforcing systemic inequalities" (Richardson, 2021, p.70). But the accuracy of the digital representation of a real-life phenomenon is dependent on the availability, accessibility, and compatibility of the data (Tzachor et al., 2022).

Emerging city scale digital twin projects like Smart Cambridge and Sidewalk Toronto are implementing smart city initiatives through the combination of data, broadband and digital twins (Oti-Sarpong et al., 2022). The combination of the data layer and digital twin transform how we live (Oti-Sarpong et al., 2022) which introduces a novel form of data-driven urbanism that uses data and algorithmic calculation for city planning. This type of approach is claimed to be fostering a better and smarter future for us while in the meantime raising concerns from scholars about privacy, human liberty and fairness with regards to bias inherent in algorithmic computation (Helbing, Fanitabasi, Giannotti, Hänggli, Hausladen, van den Hoven, Mahajan, Pedreschi & Pournaras, 2021; Helbing & S'anchez-Vaquerizo, 2022; Brauneis & Goodman, 2018; Kitchin, 2016; Juvenile Ehwi, Holmes, Maslova & Burgess, 2022; Charitonidou, 2022; Sunstein, 2022). Batty (2018) stated the issue of digital twins is that they are just replicating the real-world structure without using the real-time data or taking into consideration the real-time process that involves human daily activities. Helbing and S'anchez-Vaquerizo (2022) stated that a data-driven approach is oversimplifying human life, especially the cultural or ethical values that cannot easily be measured and quantified. It raises the concern of how data represents people's values under different circumstances. Richardson (2021) also pointed out how data is simplifying human social interaction if we heavily rely on it for solving societal issues. Several researchers are also in doubt about the efficacy of using digital twins in smart cities as the interoperability and data exchange between systems are lacking standardization (Ferré-Bigorra et al., 2022; Richardson, 2021). Further, Brauneis and Goodman (2018) are also in line with this notion and question whether the policy makers (e.g. system owners) understand how the system designs and the data affect their decision making in urban planning.

Since the application of digital twins in urban planning is still relatively young, a majority of the digital twin literature is still referencing the usage of digital twin models in the industrial sector (Ferré-Bigorra et al., 2022). Although there is a rising number of research incorporating digital twins in urban planning or smart cities, researchers have diversified focus on how to develop this model to address grand issues, how to scale it up to city level (Evangelou et al., 2022), or explore certain topics such as marginalization, and digital divide (Oti-Sarpong et al., 2022; Tzachor et al., 2022). Some researchers are focusing more on the utility aspect on how to incorporate different kinds of technologies like internet of things, artificial intelligence (AI) or sensors in a digital twin to drive smart city goals (Lee et al., 2021; Huang et al., 2022), while

others are starting to address the difficulties relating to the implementation of city scale digital twins (Batty, 2018; Ferré-Bigorra et al., 2022; Richardson, 2021).

As previously mentioned, the prevalence of algorithm calculation and data used in a city scale digital twins also raises concerns about the risk of algorithmic bias (Batty, 2018; Ferré-Bigorra et al., 2022; Madsen et al., 2022; Herath & Mittal, 2022; Richardson, 2021; Evangelou et al., 2022). Although the attention to bias in algorithmic processes is growing in AI literature, it remains "underexplored" (p.825) in the literature on digital twins (Tzachor et al., 2022). Since biases may be rooted in the design aspect of digital twins due to the intensive algorithms and data usage for smart cities, we believe that algorithmic bias may cause contradictions to at least one of the smart cities' goals, including inclusion, fair redistribution of resources and opportunities, as well as citizens' quality of life (Pierce et al., 2017). It is being argued in the literature that ethical implications in digital systems based on algorithmic calculations need to be further researched (Rai, Constantinides & Sarker, 2019; Gal, Hansen & Lee, 2022). Algorithmic calculations and data-driven innovations pose ethical risks in relation to fairness, accountability and transparency (Akter, McCarthy, Sajib, Michael, Dwivedi, D'Ambra & Shen, 2021). However, it is hard to reach a consensus regarding to what extent it is considered to be fair, transparent and accountable enough and what is considered as important values (Johnson, 2022; Buhmann, Paßmann & Fieseler, 2019).

As different socio-technical challenges involved in city scale digital twins are both complex and controversial, involving conflicts of interest, stakeholders and agendas, researchers are suggesting examining it under the ethical lens in order to understand the rationale behind each action with criticality and rationality (Juvenile Ehwi et al., 2022; Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Dubber, Pasquale, Das & Goodman, 2020; Buhmann et al., 2019; Häußler, 2021; Tzachor et al., 2022). Besides, researchers also stated that the discussion between the public and legal sphere is insufficient (Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022). Therefore, there is a lack of democratic considerations, entailing citizen engagement and legal development, which are deemed to be important aspects of city scale digital twin initiatives (Helbing et al., 2021; White et al., 2021; Zhang et al., 2022; Schiavone, Appio, Mora & Risitano, 2020; Dembski, Wössner, Letzgus, Ruddat & Yamu, 2020).

According to Mingers and Walsham (2010), when achieving democracy, it is required to fulfil pragmatic, ethical and moral domains respectively with deliberative discourse, using dialogue as engagement to reach consensus, when encountering the latest technologies and enhancing understanding of the rationale behind stakeholders' intention. In general, the pragmatic domain refers to what will be the most efficacious (Mingers & Walsham, 2010); the ethical domain refers to what will be considered as goodness and virtuous (Mingers & Walsham, 2010); and the moral domain asks the question of what will be equally beneficial to everyone (Mingers & Walsham, 2010). The democracy principle supervises the pragmatic, ethical and moral domain and governs norms that are legally established and agreed upon by all citizens through a legislative process that conforms to legal procedures (Mingers & Walsham, 2010). Hence, this thesis will utilise the principle of discourse introduced by Mingers and Walsham (2010) as an instrument to understand the decision in the data management processes of digital twins development under the smart cities initiatives as well as when encountering algorithmic bias. This thesis intends to delineate multiple facets that practitioners and urban planners may take into account while managing data and addressing algorithmic bias during the data management processes as such considerations may carry the potential risk of compromising a specific domain.

After specifying the problem area, purpose, research question and delimitations of this thesis, a second chapter will follow entailing a comprehensive literature review. The literature review will cover previous relevant papers surrounding this area of research, including algorithmic bias and the principle of discourse. Chapter 3 will describe the methodological stances of this thesis, including research philosophy, data collection and analysis methods, reliability and overall quality implications of this thesis. Chapter 4 will include the empirical findings induced from the data collection. After presenting the empirical findings, a discussion will be held contrasting the findings and their relevance in relation to existing research, after which conclusions or interesting discoveries will be presented.

1.2 Research Problem

The application of digital twins in the urban context is still emerging (Ferré-Bigorra et al., 2022), it already caught researchers' attention concerning possible negative outcomes regarding various forms of privacy violation (Kitchin, 2016). Multiple researchers also emphasized the risk of algorithmic bias due to heavy data collection and algorithms being used for driving urbanism, possibly resulting in discrimination, constraints of free will and unjust treatment (Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Brauneis & Goodman, 2018; Kitchin, 2016; Juvenile Ehwi et al., 2022; Charitonidou, 2022; Sunstein, 2022). Algorithmic bias is a socio-technical challenge requiring closer attention in the city scale digital twin setting (Tzachor et al., 2022).

As different socio-technical challenges involved in city scale digital twins are both complex and controversial, involving conflicts of interest, stakeholders and agendas, researchers are suggesting examining it under the ethical lens in order to understand the rationale behind each action with criticality and rationality (Juvenile Ehwi et al., 2022; Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Dubber et al., 2020; Buhmann et al., 2019; Häußler, 2021; Tzachor et al., 2022). Helbing et al. (2021) suggests that ethics provide directions to solve the socio-technical tensions in a deliberative way when legal systems have limitations.

Building on the aforementioned concern of algorithmic bias raised by city scale digital twins and the call for ethics scrutiny from researchers, we found that existing literature focus on using ethics to discern the general concerns in the city scale digital twin context (Ananny, 2015; Kitchin, 2016; Juvenile Ehwi et al., 2022; Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Dubber et al., 2020; Mingers & Walsham, 2010; Buhmann et al., 2019; Gal et al., 2022; Häußler, 2021) without exploring the effect of algorithmic bias. Stemming from algorithmic bias in the urban context, there is a lack of democratic considerations, entailing citizen engagement and legal development, which are deemed to be important aspects of city scale digital twin initiatives (Helbing et al., 2021; White et al., 2021; Schiavone et al., 2020; Zhang et al., 2022; Dembski et al., 2020). According to Mingers and Walsham (2010), when achieving democracy, it is required to fulfil pragmatic, ethical and moral domains respectively with deliberative discourse, using dialogue as engagement to reach consensus, when encountering the latest technologies and enhancing understanding of the rationale behind stakeholders' intention.

As mentioned previously, data-driven urbanism and algorithmic bias are underexplored in the literature. Therefore, this thesis will examine how city scale digital twin practitioners prioritize pragmatic, ethical and moral domains in the decisions throughout the data management process, such as data collection, usage, storage and sharing processes, ensuring data quality and when

encountering algorithmic bias during planning or designing city scale digital twin systems. The aim of this thesis is to outline various aspects for practitioners and urban planners to consider when handling data management processes and mitigating algorithmic bias, since it comes with a risk that a certain domain is being traded off due to various considerations.

1.3 Purpose

The focus of this thesis is to examine how city scale digital twin practitioners prioritize the democracy domains declared in the principle of discourse, namely pragmatic, ethical and moral domains (Mingers & Walsham, 2010), in various stages connected to data management, including data collection, usage, storage, sharing and quality processes in the endeavour of city scale digital twin systems, independently and in connection to algorithmic bias. Interviews with various stakeholders will provide empirical data from real life digital twin projects on how decisions are weighted and any trade-offs within the three democracy domains. This research is aiming to describe which aspects under those domains are prominent in practice and which aspects need further consideration towards deliberate city scale digital twin models, sequentially mitigating bias in the endeavour. The results of the thesis can act as a guidance and a point of departure for public and private sectors' further democratic discussions regarding digital twins in the urban settings, possibly aiding policy makers and reminding stakeholders of their accountability in relation to city scale digital twins.

1.4 Research Question

> How are pragmatic, ethical, and moral considerations prioritized by digital twin practitioners during data management processes, including collection, usage, storage, sharing and quality as well as when encountering algorithmic bias in those processes?

1.5 Delimitation

The aim of this thesis is not to focus on the difficulties of developing digital twins for smart cities, but rather to dig into the development process in terms of human decisions and considerations in the data management process and in relation to algorithmic bias. The goal is not to define the rights and wrongs in this endeavour but rather describe how the democracy domains within the principle of discourse are represented in the discussions on planning and designing city scale digital twins. It is further outside of the scope of this research to go into depth regarding the algorithms used for processing data in the digital twin projects.

Furthermore, although we touch on the topic of policy in the context, we are not aiming to produce enough research depth on the legal perspective to draw conclusions within that area of research, as this thesis will focus on the socio-technical perspective in the information systems field. Rather, the conclusions in this thesis might provide a useful basis for further development of policies within the city scale digital twin area.

2 Literature Review

2.1 The Concept of Smart Cities

Smart cities have been of significant prominence in urban development discussions, where it has been coupled with concerns for growing environmental, socioeconomic, and governance challenges stemming from rapid urbanization (Yigitcanlar et al., 2018). The smart city concept originated in the smart growth movement (Mohammadi & Taylor, 2018; Yigitcanlar et al., 2018; Dembski et al., 2020) which is dedicated to the three E's of sustainability: economy, equity, and environment (Wu, Yin, Zhou & Ye, 2017; Mohammed, Alshuwaikhat & Adenle, 2016). Smart growth was a response to unlimited urban growth, trying to promote innovative thinking to solve issues like environmental degradation, decreasing ecosystems, transportation problems, and uneven distribution of benefits from economic and environmental conditions (Mohammed et al., 2016). The purpose of the smart city is to develop a city in a sustainable manner, from a cultural, socio-economic, and environmental perspective (Schiavone et al., 2020). A successful smart city integrates people, systems infrastructure and technology to cultivate city performance (Mohammadi & Taylor, 2018) with an aim to increase efficiency of city infrastructure and citizen's services (Azevedo Guedes, Carvalho Alvarenga, dos Santos Sgarbi Goulart, Rodriguez y Rodriguez & Pereira Soares, 2018; Zhang et al., 2022).

The concept of a smart city is explained by Pierce et al. (2017) to include three conceptual layers. An environment layer that produces data, the data is contained in a digital layer which can be utilized in the city's service layer, to catalyse smart innovation and innovative solutions in order to reach the smart city goals (Pierce et al., 2017). The environment layer contains everything that makes up the physical space; terrain, buildings, infrastructure such as transport systems, and mobility entailing peoples movements and behavioural routes (White et al., 2021), as well as socio-cultural beliefs, laws, and practices (Pierce et al., 2017). Data stemming from the physical space, from citizens, devices and sensors, are collected in the digital layer (Pierce et al., 2017; White et al., 2021). The service layer, including idea generation and development, creation and maintenance, analysis, and governance (Pierce et al., 2017), uses data from the digital layer to monitor and manage a myriad of societal functions, such as transportation, waste management and crime detection (White et al., 2021), with the help of technology (Azevedo Guedes et al., 2018). The combination of a digital layer and a service layer has also been referred to as a smart city layer, and its result is iteratively fed back to the environment layer (White et al., 2021).

Stemming from a review of 110 selected research papers, the most important drivers of smart city development around the world have been found to include urban planning, city infrastructure, mobility, public safety, health, sustainability, public policies and urban risks (Azevedo Guedes et al., 2018). It is further argued that technology should be utilized as means to an end, and not as an end in itself (Yigitcanlar et al., 2018). The technological innovations in the city should in other words work towards the goals of making the city smarter in the first place. The goals of a smart city contain nuances, nevertheless, an overarching description can be found in the framework of Pierce et al. (2017), where the goals include increasing city resilience, sustainability and quality of life. These goals are partly or fully confirmed by multiple researchers (eg. Mohammadi & Taylor, 2018; Yigitcanlar et al., 2018; Tzachor et al., 2022; Azevedo Guedes et al., 2018). Yigitcanlar et al. (2018) the goals of a smart city contain nuances, nevertheless, an overarching description can be found in the framework of Pierce et

al. (2017), where the goals include increasing city resilience, sustainability and quality of life. These goals are partly or fully confirmed by multiple researchers (eg. Mohammadi & Taylor, 2018; Yigitcanlar et al., 2018; Tzachor et al., 2022; Azevedo Guedes et al., 2018). Yigitcanlar et al. (2018) states that goals and desired outcomes of the smart city should be initially clarified.

By way of illustration, see the parking API of Copenhagen Solutions Lab (2018). The API collects data from existing parking applications, stemming from peoples parking behaviour, to calculate and predict available parking spaces (Copenhagen Solutions Lab, 2018). The API can reversely be implemented in said parking applications, or GPS systems to provide easier parking for the citizens and reduce time spent unnecessarily driving around in the city (Copenhagen Solutions Lab, 2018). The API is working towards goals such as quality of life and reduced carbon emissions. To manage increasing complexities as cities grow, paired with an increasing desire to develop sustainable cities, a digital twin can be utilized as a smart innovation in the city setting (Mohammadi & Taylor, 2018). Digital twin creation and prediction can act as basis for decision-making in the endeavour of gaining a deeper understanding of how the social environment is interacting with the physical environment (Mohammadi & Taylor, 2018).

2.2 City Scale Digital Twins

The digital twin originated from the manufacturing industry, with the purpose of improving efficiency in operations (Ferré-Bigorra et al., 2022; Tzachor et al., 2022) but is being increasingly utilized on an urban scale (Evangelou et al., 2022; Lee et al., 2021; Huang et al., 2022; Herath & Mittal, 2022). The digital twin is defined as a virtual replica of the physical space, in this case replicating the structure of both living and dead entities in a city (Tzachor et al., 2022; White et al., 2021; Dulaimi et al., 2022). Data collected in the digital (smart city) layer with the integration of IoT, cloud computing, big data, AI and more (Deren et al., 2021), is being utilized in a virtual (digital twin) layer, to do simulations and predictions (White et al., 2021; Tzachor et al., 2022). Data transfers from the physical space to the virtual space, gets processed and is transformed into information flowing back into the city layers (Deren et al., 2021; White et al., 2021). Deren et al. (2021) refers to smart cities based on digital twins as digital twin cities, which they state consist of four main characteristics: accurate mapping, virtual-real interaction, software definition, and intelligent feedback. These characteristics hold through further literature, however represented with differentiated taxonomy or less explicitly (eg. Zhang et al., 2022).

Accurate mapping has the purpose to "fully perceive and dynamically monitor the city's operating status" (p.2), using sensors and other devices to collect data about all aspects of the physical space (Deren et al., 2021). Geographic information systems (GIS), building information modelling (BIM), IoT and other technologies provide potential for "multidimensional data aggregation" (p.112) to create profound coverage of the city complexity, including both static and dynamic entities such as roads, buildings, and population or traffic flow (Zhang et al., 2022). *Virtual-real interaction* entails the ability to search for physical traces of "people, logistics and vehicles" (p.3) in the virtual replica once generated (Deren et al., 2021). The physical space is "enriched, extended and expanded" (p.112) in the virtual space where managers and decision-makers might interact with the virtual replica to search for, modify, analyse and conduct simulations on physical city entities (Zhang et al., 2022). *Software definition* is the concept of using software platforms to replicate the physical space and make simulations regarding people, situations and entities in the virtual space (Deren

et al., 2021). Conducting analysis from IoT intelligence, spatial geography and urban data, to gain insight on risk, challenges, and strategies with the purpose of improving urban operations (Zhang et al., 2022). *Intelligent feedback* is the ability to identify early warning signs through the process of planning, designing and simulating the virtual replica of the city, the feedback is aimed at providing countermeasures when identifying potential risks, conflicts and dangers (Deren et al., 2021). The platform can predict risks, present the city's current status along with suitable countermeasures that can easily be deployed by managers and decision-makers to protect citizens lives and property (Zhang et al., 2022).

Application areas of a digital twin in the urban setting are indicated throughout the literature. Traffic management (White et al., 2021), smart energy grids (Dembski et al., 2020), efficient resource allocation (Tzachor et al., 2022), and public services (Zhang et al., 2022) to name a few. Digital twins in a city can contribute to optimization of planning and management, catalysing improvement of citizen services and construction of the urban spaces (Deren et al., 2021). Collecting data about people's movements in urban spaces and throughout the city infrastructure that can be transformed into information about groupings and patterns of people's behaviour and opinions, can sequentially aid city planners and policymakers to improve services along with the physical infrastructure in a manner that better meets the needs of the citizens (Mohammadi & Taylor, 2018) in a shorter amount of time (Zhang et al., 2022).

> Smart cities based on digital twins have broad prospects for economic transformation, urban smart management and public smart services, so that man and nature can develop more coordinately. (Deren et al., 2021, p.10)

Furthermore, "[s]takeholders are important components of digital twin city ecology, including citizens, governments, designers, builders and many other subjects" (Zhang et al., 2022, p.114). A digital twin can serve as a tool to improve collaborative planning processes (Dembski et al., 2020), to efficiently coordinate multiple stakeholders and complex technologies long-term (Zhang et al., 2022). In the smart city setting coordination between stakeholders and differentiated actors is necessary to promote a more efficient smartification of cities (Schiavone et al., 2020). Moreover, White et al. (2021) points towards citizen feedback as a strategy to move beyond the initial options and further aid the responsible units during city planning, as it adds another layer of data. Dembski et al. (2020) also put emphasis on citizen participation, pushing to consider citizens' experiences, including marginalized groups. This is a view shared by Mohammadi and Taylor (2018) as well as Zhang et al. (2022) who similarly call for a "people-oriented and more inclusive" (p.114) approach to city digital twin development.

The reliability and accuracy of the city's virtual representation, depends on the extent to which there is available, accessible and compatible data (Tzachor et al., 2022). The virtual representation of the physical space is important since it provides basis for decision-making within the city, and although the issue of data acquisition can be mitigated with improved stakeholder collaboration (Tzachor et al., 2022), data related challenges remains. When utilizing a digital twin in the urban sphere, the complexity of a city's structure, relationships and activities, poses additional challenges such as data handling, interoperability, processing, and interpretation (Shahat, Hyun & Yeom, 2021). While the goal is not to present a perfect digital twin on the first try, but rather to improve the model with additional data incrementally (White et al., 2021), this leads us into the data phenomenon within the city scale digital twin context.

2.2.1 *Data-driven Urbanism*

In connection with smart city goals, they are derived from urban issues or challenges in reality and anticipated to achieve the goals with the use of ICT to simulate and predict situations for practicable solutions (Brauneis & Goodman, 2018). Digital twins are one of those technologies exploited to serve those purposes. Different layers of digital twins consist of a significant part of urban data including geospatial, people movement and real-time operational data of the city activities such as traffic, public utilities or services (White et al., 2021). Those data are collected, analysed and modelled for decisions regarding urban planning, policy making (White et al., 2021), as well as efficiently coordinating and responding to emergencies (Deren et al., 2021). By leveraging vast amounts of urban data and utilizing city analytical tools, these approaches are paving the way in shaping and dictating urban agendas and exerting influence, control and governance over how city systems respond and operate as a novel form of data-driven urbanism (Kitchin, 2016). Dubber et al. (2020) stated that the flow of data is leading people to select greater options spanning across planning, maximizing efficiency or exercising control as people believe that data represents facts or scientific evidence (Brauneis & Goodman, 2018). However, Helbing et al. (2021) questioned whether big data and algorithmic calculation in an urban setting can straightforwardly take into account human choices and behaviour as a business setting to generate the most optimal option, where cities have different agendas and conflicts of interest that are not easily aligned, as well as where goals may run parallelly. This notion resonated with (Nochta, Wan, Schooling & Parlikad, 2020) that different backgrounds of people, mindsets, interest and agendas created the complexities in forming consensus or decisions towards urban policy and planning. Regarding data-driven urbanism, researchers raised the questions and concerns towards data-driven urbanism alongside with the possible negative outcome regarding confidentiality, discrimination, liberty and unjust treatment (Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Brauneis & Goodman, 2018; Kitchin, 2016; Juvenile Ehwi et al., 2022; Charitonidou, 2022; Sunstein, 2022). Those concerns are manifold throughout the processes from the collection, storage, usage and sharing as well as its quality.

2.2.1.1 *Data Collection*

Unlike data informed urbanism, data is no longer collected for a single purpose or limited in scale that only shows the "snapshot of cities at particular moment" in aggregated level, but being gathered in full scale and in real-time through the employment of IoT and ubiquitous sensing which contribute to higher level of data granularity (Kitchin, 2016, p.2). The surveillance sensors are everywhere to capture individual data in conjunction with urban data to oversee and regulate the various aspects and dimensions of the city at all levels and magnitudes that arise the privacy issue (Helbing et al., 2021; Brauneis & Goodman, 2018; Petrova-Antonova & Ilieva, 2021; Kitchin, 2016; Juvenile Ehwi et al., 2022; Charitonidou, 2022). The data are retrieved in real time or near real time manner with automatic connectivity and shared with different siloed systems that are developed by third parties for the sake of smart cities' purposes that creates difficulties to notice or consent affected individuals due to its data volume and opacity (Dubber et al., 2020). Even though personal sensitive data could be anonymous through processing, Kitchin (2016) argued that the tracking data, especially the people movement data showing the pattern of how frequent a person visits certain places may still produce inferences from algorithmic prediction that lead to "predictive privacy harm" such as revealing one's sexual orientation or certain interests and may even be inaccurately predicted resulting biases or discrimination (p.8). Papyshev and Yarime (2021) pointed that no standard procedure or guidance in data collection causing the missing linkage between structured and

unstructured data, missing data values or certain type of data which is compromising the representation of reality and city scale digital twins' functionality (Ferré-Bigorra et al., 2022).

2.2.1.2   Data Storage & Use

In order to process massive amount of gathered urban data in city scale digital twins, algorithms are heavily in use to cater the data interoperability which may inherit biases or discrimination from human judgment consciously or unconsciously and lead to unjust decisions made towards affected groups of people as well as externalities (Helbing et al., 2021; Dubber et al., 2020; Brauneis & Goodman, 2018; Kitchin, 2016; Charitonidou, 2022; Stinson, 2022). Hämäläinen (2021) posed another view meaning that the scenarios created in city scale digital twins are based on actual data that is representing the reality and removing the human sentimental decision or biased judgment from it. This view is echoed to a point drawing from Brauneis and Goodman (2018) article that individuals have a tendency to trust in data as scientific truth, but the author also pointed out the flaw from this view that data may be biased due to various reasons.

Furthermore, a data-driven society is oversimplifying human's nature and creating "an inhuman society" as human desire and values are interpreted as data blocks (Helbing & S'anchez-Vaquerizo, 2022, p.9). Data can tell the facts of certain situations but it may miss out other societal aspects and human values like unity, shared values and norms that are hard to be measurable or comprehended (Helbing & S'anchez-Vaquerizo, 2022). To a large degree, human interaction and their pattern of living is being neglected in digital twins settings (Helbing & S'anchez-Vaquerizo, 2022); Dubber et al. (2020) emphasized the fact that "technological determinism" in smart cities is nudging humans to adjust their existing living style to adapt the changes created by technologies in expense of their free will with the priority to put city planning ahead of policies (p.5). After all, the ideal of data-driven urbanism is that everyone goes online so that data can be interoperable in the city setting (Charitonidou, 2022). Charitonidou (2022) addressed the myth of "data universalism" that digital divide will still exist in city scale digital twin contexts as inaccessibility of data and technologies are still happening in real life and certain groups of people's rights are under provision (p.238).

Other than the data usage, the uncertainty of data volume leads to the concern about data storage and its security (Nochta et al., 2020). Dubber et al. (2020) stated that the city authority should keep abreast of the role as a guardian to protect the data of the public is being exploited by the private organizations and in the meantime, control and balance the use of data by cooperating with the private companies to make good use of the data potential and capabilities for the society. The author also supplemented the concern that certain private institutions obtaining preferential technological assets like massive amounts of urban data and algorithms may exercise their power influencing urban planning for their own sake and cause detrimental effects to neoliberal society, "a society's political and economic institutions should be robustly liberal and capitalist, but supplemented by a constitutionally limited democracy and a modest welfare state" (Vallier, 2022). However, researchers also agreed that the implementation of city scale digital twins is costly in comparison to other city monitoring and governance technology approaches due to high computational capacity and storage is required for data (Ferré-Bigorra et al., 2022; Papyshev & Yarime, 2021) which consequently may compromise in terms of data quality in respective to money concern (Ferré-Bigorra et al., 2022).

2.2.1.3   Data Sharing

Dubber et al. (2020) stated that by opening the urban data access as common resources can prevent privatization of data by certain parties to extract the value out of it without returning it back to the society and less motivated to over collect data for its own use. The author also stated the notion that private entities may manipulate the city planning decision based on how much urban data that they have owned. While Helbing et al. (2021) proposed that the source of data is coming from the general public, the ownership and custodian should be the people themselves and they have the right to determine who they want to share the data with in an aggregated or detailed level. The data ownership is hard to define especially for those data gathered from the public through sensors. In addition to the mentioned data source, some urban or operational data is shared by the public sector or procured from 3rd parties like private companies (Ferré-Bigorra et al., 2022). Those parties are less motivated to share their data (Papyshev & Yarime, 2021) due to the lack of commercial interest, data ownership and sharing framework (Nochta et al., 2020), business concern such as the risk of leaking their business secrecy or customer's privacy (Leese, 2014; Ananny & Crawford, 2018; Burrell, 2016; Buhmann et al., 2019). Also, the justification and clarity of possible benefits about the value of data collection, sharing and collaboration with different societal stakeholders is insufficient which created resistance to sharing data for building city scale digital twin purposes (Nochta et al., 2020).

2.2.1.4   Data Quality

In regard to the aforementioned issues, Ferré-Bigorra et al. (2022) stated that data quality is problematic as it is unevenly distributed, skewed or mistaken by humans, especially that data coming from individual mobile devices' sensors are in low precision and accuracy level. Furthermore, the dilemma between cost and worthy versus data accuracy regarding data collection, storage, usage and sharing is another factor causing variation in data quality (Ferré-Bigorra et al., 2022). For instance, the frequency of gathering data for city landscapes is relatively less cost efficient and worthy than those from urban operational systems as the activities are less likely to change from the spatial temporal perspective (Ferré-Bigorra et al., 2022). Moreover, various data standards are on the market for three dimensional city modeling such as geospatial data which has its own standards like CityGML, Open Geospatial Consortium or Geographic Data Files but no common standard is defined which is hampering people to maintain the data quality like its metadata and data versioning as it is all dependent to human judgment and their decisions based on different circumstances (Papyshev & Yarime, 2021).

2.3 Algorithmic Bias

The application of city scale digital twins are still in an exploratory stage (Zhang et al., 2022), with its complex relations and activities posing further challenges compared to its predecessor in industry (Shahat et al., 2021). From a perspective of data-driven urbanism, bias risks occurring in several stages related to data; collection, preparation, modelling and evaluation (Akter et al., 2021). Being in an exploratory stage entails challenges of its own, for example immature technology and insufficient conceptual understanding (Zhang et al., 2022), but although algorithmic bias is underexplored in the digital twin literature, several issues have been identified in AI literature (Tzachor et al., 2022). Examples include misrepresenting minorities, racism and sexism, undermining a myriad of the UN's social sustainability targets,

in *The 2030 Agenda for Sustainable Development* (Tzachor et al., 2022). Using historical decision data to train systems risk embedding historical discrimination into future decision making, having uncovered cases within policing, hiring, medical issues and more (Stinson, 2022). Scholars are increasingly cautious of algorithmic bias emerging from multiple origins and in any development stage in relation to data-driven innovations (Akter et al., 2021).

Kordzadeh and Ghasemaghaei (2022, p.395) define algorithmic bias as a "systematic deviation from equality that emerges in the outputs of an algorithm". It occurs when discriminatory results are produced from inaccurate or missing relations in the modelling of the system (Akter et al., 2021). Scholars are concerned that societies' existing biases risk being amplified if those biases permeate training data and the algorithms themselves (Rai et al., 2019). Relevant risks of bias in the city setting include limited accessibility for vulnerable groups and people when it comes to resources or pricing (Akter et al., 2021). Algorithmic bias disproportionately affects and discriminates against already marginalized groups and people (Stinson, 2022; Kordzadeh & Ghasemaghaei, 2022). Biases in data-driven innovation risk producing "unjust, unfair, or prejudicial" (p.1) results based on gender, race, sexual orientation and more (Akter et al., 2021). Illustratively, Stinson (2022) gives the example of police in the US stopping black men at a disproportional rate without a cause, a higher frequency of "petty" crimes will be identified among black men compared to other groups because of bias within the police force which will lead to continuous disproportional police activity connected to black men, reinforcing the discrimination.

There is some consensus about where algorithmic bias stem from. Akter et al. (2021) identify three categories of algorithmic bias from a literature review analysing 40 scientific papers: societal bias, data bias, and method bias, Stinson (2022) identifies corresponding categories, biased people, including societal bias, biased data, and biased algorithms, including method bias. A biased society, inhabiting people having bias is the first problem in this regard. It is people that build algorithms and collect data, embedding their own biases in the systems (Akter et al., 2021; Stinson, 2022). Stinson (2022) state a need to further include minority groups in development processes, since the negative outcomes usually stems from developers' lack of perspective and knowledge. Illustratively, Akter et al. (2021) discusses an Australian case of using machine learning to determine social security payments for citizens with unstable financial status, the system was designed in a manner that in multiple ways made it more difficult for people with limited resources to receive support from the government, stemming from a limited perspective during development, making life harder for already vulnerable people in society. Bias in data can arise for several reasons. Developers were historically homogenic, looking for data from sources they were already familiar with, in other words lack of representation and diversity can lead to biased datasets (Stinson, 2022). This type of non-random selection of data points is called selection bias (Akter et al., 2021; Stinson, 2022). Illustratively, Akter et al. (2021) explains an example where Amazon stopped using an AI-based recruitment system, the training data was lacking female applicants leading to disproportional recruitment of male applicants. Skewed training data is one of the primary factors leading to bias in AI (Zou & Schiebinger, 2018). Another aspect of algorithmic bias is biased algorithms, which result in phenomena like homogenization, over-specialization (Stinson, 2022), and correlation fallacy (Akter et al., 2021). The bias might furthermore reinforce over time (Kordzadeh & Ghasemaghaei, 2022), for example as the algorithm optimizes its accuracy leading to disproportionately accurate predictions for the majority group in contrast to optimizing softer values such as user satisfaction (Stinson, 2022). Lack of explainability and testability of the algorithm predictions might further cause confirmation bias, meaning people's biases are confirmed in the outcome of the algorithm and not questioned

properly (Akter et al., 2021). Akter et al. (2021) stress the need to effectively handle all three aspects: people, data, and algorithms, to create fair data-driven innovations.

In the endeavour to mitigate bias, Akter et al. (2021) emphasizes citizen feedback and stakeholder engagement, to produce regular insight about datasets to create fairer systems that better meets the needs of the customers, or in this case, citizens. Zou and Schiebinger (2018) recommend incorporating standardized metadata into development processes to help identify and mitigate sources of bias. The evaluation bias might include a variety of metrics. Individual-level metrics try to ensure individuals with similar qualifications in a specific context will have the possibility to produce similar outcomes, while group-level metrics have the purpose of counteracting disproportional negative affects towards certain groups of people (Kordzadeh & Ghasemaghaei, 2022). Illustratively, Kordzadeh and Ghasemaghaei (2022) explain evaluation of bias in a loan application context. The authors propose grouping the dataset by gender and sequentially produce a confusion matrix of the results, including true positives, true negatives, false positives, and false negatives. Since gender should be an irrelevant metrics (irrelevant metrics are proposed and protected within the law) in this context, the confusion matrix should not present false negatives disproportionally related to any gender, to be considered unbiased. However, metrics and bias evaluation methods differ for each new case, other methods include demographic parity, equal opportunity or equalized odds, balanced error rates and so on (Kordzadeh & Ghasemaghaei, 2022). As algorithmic bias has ethical implications connected to fairness, accountability and transparency (Akter et al., 2021), researchers are calling for further investigation in relation to decision making surrounding these systems (Johnson, 2022; Buhmann et al., 2019).

2.3.1 *Fairness, Transparency & Accountability*

AI versatile capabilities are serving different tasks and purposes especially in algorithmic decision making (Johnson, 2022). Biases can be consciously or unconsciously built in algorithms through various means (Johnson, 2022; DeVos, Dhabalia, Shen, Holstein & Eslami, 2022; Stinson, 2022). Under the context of city scale digital twins, algorithms and data are extensively being used in order to achieve goals like resource optimization, controlling and monitoring (Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Dubber et al., 2020). This kind of approach raises certain concerns or negative impacts like privacy, equality and digital divide (Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Dubber et al., 2020; Buhmann et al., 2019) that are getting scholars' attention. Thus, researchers are calling for examining fairness, transparency and accountability towards artificial intelligence (AI) or any computerized decision making through algorithms (Johnson, 2022; Buhmann et al., 2019).

Fairness refers to everyone who should be treated equally. When it comes to the context of algorithm fairness, the decision made from the computational calculation should not introduce any negative consequences like discrimination or unjust treatment to any individuals or demographic group (Yang & Stoyanovich, 2017). Similarly, transparency is considered as the openness and clarity by sharing every information explicitly to the concerned parties. The revealed information should be understood by the stakeholders and can be used by them when necessary (Johnson, 2022). However, the algorithms being used in AI are in nature defined as a black box (Mahmud, Islam, Ahmed & Smolander, 2022; Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Stinson, 2022; DeVos et al., 2022) as its development is fluid and evolving that gains in size (Buhmann et al., 2019). The opacity in algorithms raises the concern that whether people who rely on algorithmic decision making are aware of its shortcomings and who will be taking accountability of its consequences (Buhmann et al., 2019; Stinson, 2022).

When the algorithms reinforce learning itself and biases are inherited from diversified sources, it comes with the question that who can be responsible for the end result if they have limited knowledge about it.

The solution to the algorithmic opacity is transparency (Kroll, Huey, Barocas, Felten, Reidenberg, Robinson & Yu, 2017). Johnson (2022) proposed that the more information provided to enhance the transparency in algorithms, people are less prone to develop biased and dangerous algorithms. This statement is half true as Stinson (2022) stated that certain biases can be unconsciously built in algorithms such as statistical bias or people bias from their moral sense of viewpoint. Explainability is another equivalent concept of transparency and being strengthened by the European Union General Data Protection Regulation (EU GDPR) (Johnson, 2022) that individuals have the right to request explanation on the reasoning behind the automated decision making or the importance and expected outcomes of themselves when their data are being processed (Intersoft Consulting Services AG, 2016). According to Johnson (2022), the complexity of algorithms is hindering the general public or even the experts to fully understand its underlying logic in a very detailed level. Also, it is difficult to define to what extent it is considered to be an acceptable level of explaining the causation of its outcome. Brauneis and Goodman (2018) filed multiple open record requests to different types of entities that built predictive algorithms for "criminal justice systems or civil systems" to disclose their information regarding the algorithm's coding and its development process under the state freedom of information laws (p.109). The results showed that the quality of information varied in different levels from simply revealing basic algorithms to more complicated algorithms and even its development processes. Kroll et al. (2017) pointed that full transparency is not possible in the real world as the decisions to disclose information intervening legal or business consideration. How much explainable information can we dig into and how much information is willingly to be divulged under the name of transparency? Even under the designation of law, the amount of information being disclosed is by choice and selectively. Increasing the accessibility to information does not always improve transparency as it is also dependent on the meaningfulness of the provided information to the audiences (Brauneis & Goodman, 2018). Likewise, providing massive amounts of information like algorithms to the general public will only bring back opacity if the audiences lack the technical knowledge to interpret the technical rules and probabilities (Buhmann et al., 2019; Kroll et al., 2017).

In view of the strain in achieving transparency, Kroll et al. (2017) suggested that transparency is only a partial solution to accountability when it comes to the context of automated decision making . Johnson (2022) and Buhmann et al. (2019) stated that algorithmic accountability would be a substitute for transparency as the comprehensivity level of the algorithm does not impact its accountability. However, there are different interpretations towards accountability in terms of different worldviews that make it difficult to apply in algorithmic calculation. In the commercial world, business accounts for protecting its algorithms or the related data as a secrecy as it helps them to maintain their competitive advantages or due to policy restrictions which are considered as a "strategic accountability" (Leese, 2014; Ananny & Crawford, 2018; Buhmann et al., 2019, p.266; Burrell, 2016). While in the technical development world, "technical accountability" exists to the people who develop the algorithm have to understand what they are building and whether the data used in training the model contains biases (Burrell, 2016; Ananny, 2015; Buhmann et al., 2019). In the legal world, accountability is expressed as "a relationship between an actor and a forum, in which the actor has an obligation to explain and to justify his or her conduct, the forum can pose questions and pass judgment, and the actor may face consequences." (Bovens, 2007, p.450). Its root can be traced back to the 10th century where the sovereign is not about holding their subject to account but on the contrary, to be

answerable and responsive to their citizens (Bovens, 2007). This practice becomes an instrumental act to enhance governance under different scrutiny levels and circumstances (Bovens, 2007). Various interpretations of accountability are deemed to be reasonable, but it is still leaving the accountability question stranded with whom, what and why to bear the accountability.

Ananny and Crawford (2018) shed light on the relationship between transparency and accountability from the view that enhancing visibility to a phenomenon introduces avenues and duties to establish mechanisms for holding it responsible which results in changes. The more important implication is that rather than focusing on dissecting the algorithms, it is preferable to consider comprehensively on the socio-technical aspect that the algorithm does not possess complexity on its own, but rather generates complexity through its interconnection and interaction with both human and non-human components (Ananny & Crawford, 2018). Johnson (2022) also supported this view for examining algorithmic decision making with the interaction between the technical artifact and human being. She believed algorithms are the technological artifact instrumented for the outcome. However, the outcome cannot be achieved if there is no efficacy of the technical artifact. Additionally, a technical artifact is only efficacious when it is combined with human behaviour. In general, researchers put the spotlight on the technical artifacts such as algorithms, their usages or outcomes instead of exploring the ontological construction of the technical artifacts where humans are the main facilitators that infiltrate their decisions into those formations (Konaté, Zaraté, Gueye & Camilleri, 2020).

As previously mentioned, biases are everywhere as human decisions contain their own interpretations of the issues and intertwine their norms, cultures or opinions with their choices (Stinson, 2022). Therefore, Johnson (2022) and Buhmann et al. (2019) suggested that the future research should focus on the rationale behind the technical artifacts development. Especially in the city scale setting of digital twins, the essence of it are the smart technologies with vast amounts of algorithmic calculations and data-driven processes in various complex interoperable and silo systems (Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Brauneis & Goodman, 2018; Dubber et al., 2020; DeVos et al., 2022; Johnson, 2022; Charitonidou, 2022). In order to further explain and investigate the socio-technical relationship simultaneously, certain ethical principles and discourse ethics are in use to make algorithmic accountability in a pragmatic way under the context of either smart cities or digital twins (Mingers & Walsham, 2010; Gal et al., 2022; Dubber et al., 2020; Juvenile Ehwi et al., 2022; Kitchin, 2016; Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022).

Even though the scholars begin to derive the risks or possible impacts of the blackboxed or biased algorithms in this setting through the ethical lens (Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Dubber et al., 2020; Buhmann et al., 2019; Johnson, 2022; Stinson, 2022; DeVos et al., 2022), the human decision making towards city scale digital twins throughout the entire processes are still underexplored. Researchers particularly stated that the problematic part of using data and algorithms to drive city planning or support human decision making introduce tensions between different stakeholders, amplify ethical challenges and increase conflict of interest (Charitonidou, 2022; Brauneis & Goodman, 2018; Helbing & S'anchez-Vaquerizo, 2022). Some of the researchers also stated that the discussion between the public and legal sphere are insufficient under the current state of affairs (Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022). Hence, we believe the principle of discourse introduced by Mingers and Walsham (2010) with the combination of pragmatic, ethical, moral aspects under democracy principles and the help of the theory of communicative action will be the

instrument to understand human choices in different stages of digital twins development under the smart cities initiatives.

2.4 Principle of Discourse

The prevalence of technologies being used in individual daily life and even extended to city planning for the society introduces the debates whether it is beneficial or detrimental to our future. The balance between different contrasting agendas like economic growth, human wellbeing or grand challenges become prominent (Kitchin, 2016; Juvenile Ehwi et al., 2022; Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Dubber et al., 2020; Charitonidou, 2022; White et al., 2021). In the academia, scholars suggests the use of ethics to analyse this type of controversial topic in order to draw different dimensions of views with rational and critical consideration (Ananny, 2015; Kitchin, 2016; Juvenile Ehwi et al., 2022; Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Dubber et al., 2020; Mingers & Walsham, 2010; Buhmann et al., 2019; Gal et al., 2022; Häußler, 2021). Within the information systems area, computer and information ethics are two disciplines making an effort to address sceptical views towards technologies emerging issues.

Computer Ethics (CE) refers to certain codes of conduct for computing professionals to make decisions in conjunction with computer activities based on ethical considerations (Floridi, 1998). Since CE are trying to generalize similar real world cases in an action-oriented theoretical setting and come up with the ethical code that can be applicable to most of the issues, it leads to limited insights to be produced when new issues arise on top of the existing issues or even extend it to a broader sense (Floridi, 1998). Besides, CE neglects the human influences that are incapable of solving problems with traditional ethical principles such as consequentialism, contractualism and deontology (Floridi, 1998). While information ethics (IE) is considered as the interrelated approach to CE as all information existed in the *Infosphere* encompasses all forms of information, regardless of their physical instantiation to support moral assertion (Floridi, 1998). IE seems to complement the absence of morality and humanity by treating information as the object to inform consideration and implementation of ethical principles in real life situations. However, Stahl (2008b) considered IE to be an impractical approach with reference to Floridi's examples on linking information and its accessibility in conjunction with IE and privacy issues; which turned out to be another universal approach that reflects on the nature of existence. The result overlooks various cultural contexts in privacy as well as the intractable issues that arise with whom to account for the responsibilities. Therefore, the author concluded that IE can only provide a conceptualized ground for ethical challenges emerging in information systems or technological areas (Stahl, 2008b). In addition to that, Stahl (2008b) and Mingers (2011) suggested discourse ethics would be an alternate approach to CE and IE since it is being used in other real life business and organizational settings that exhibit its pragmatic value and its potential in other areas.

Discourse ethics (DE) is treated as a practical approach as the analysis is based on the discussion and negotiation in real life among those being impacted by decisions and further to its potential to pragmatize abstract theories through methodologies (Mingers & Walsham, 2010) which discern it from other fields of ethics like CE and IE. DE stemmed from the Theory of Communicative Action (TCA) introduced by Habermas respectively in 1984 and 1987 (Bohman, James & Rehg, 2017). Habermas emphasized that TCA's goal is to achieve consensus through communication in which everyone has the equal right to participate in the

debate and nobody can exert power over others' right to make an utterance (Mingers & Walsham, 2010).

> "The point of this process is to try to generate a common will and not just an accommodation of interests. That is, the participants should become convinced that it is genuinely the best way for all of them to resolve their common differences. To this end, (1) the mention of interests and value orientations refers to the participants' concerns within the pragmatic and ethical domains respectively; (2) participants should try and genuinely take on the perspectives and roles of the other, and be prepared to modify their own; and (3) agreement should be based, as always, on force of argument rather than force of power" (Mingers & Walsham, 2010, p.11).

In light of the complexity in city scale digital twins, we can see the conflict of interest in the smart cities goals and city planning as well as the ethical issues arising from data-driven urbanism and the risk of algorithmic biases throughout the development processes (Kitchin, 2016; Juvenile Ehwi et al., 2022; Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Dubber et al., 2020; Charitonidou, 2022; White et al., 2021). Various known or unknown stakeholders such as city planners, technical developers and citizens are involved or being impacted in this setting. TCA as the process of dialogue and justification helps to ensure that decisions are based on reasoning rather than power or self-interest, and that all stakeholders are held accountable for their contributions to the decision-making process (Mingers & Walsham, 2010). As Helbing and S'anchez-Vaquerizo (2022) stated the notion that humans make decisions based on certain levels of heuristics instead of clearly explainable by numbers, they may choose a "good enough" option instead of the most optimal one according to their different roles and goals in specific situations and change over time with the influence of their norms and culture (p.8). Those human determinations would be as of the research interest to be further scrutinized under DE's pragmatic, ethical and moral domain with the question starting from *how ought we do* (Mingers & Walsham, 2010).

In general, the pragmatic domain refers to what will be the most efficacious (Mingers & Walsham, 2010). It means to discover the most appropriate approach to successfully and effectively achieve the desired outcome with rationality (Mingers & Walsham, 2010). Consequentialism, to be understood as the action taken "maximize the overall good or minimize the overall harm", is one the ethical theories being categorized under pragmatic domain (Mingers & Walsham, 2010, p.834). The other one would be utilitarianism, "to maximize the greatest good for the greatest number and thereby risk injustice for the minority" (Mingers & Walsham, 2010, p.835). The ethical domain refers to what will be considered as goodness and virtuous (Mingers & Walsham, 2010). Virtue ethics is categorized in this domain as it represents individuals' values that are considered virtuous such as integrity and fairness as well as Aristotelian and communitarian that to build a good life among the society (Mingers & Walsham, 2010). The moral domain asks the question of what will be equally beneficial to everyone (Mingers & Walsham, 2010). Privacy is one of those moral value lying in the moral domain (DeCew, 2018). Another ethics theory is deontology, which is categorized under the moral domain as it expresses the action or duty which is considered as right or wrong, fair or unfair (Mingers & Walsham, 2010). The moral norm must be universalized, such as a universal law with justification and applicability to all the affected parties with consensus gained. Even though the ethical and the moral domains are separately indicted in the principle of discourse, the relationship between them can be explained as "morality as the set of factual social rules or

norms that accepted within a certain group or society whereas ethics is the reflection and justification of such rules or norms" (Stahl, 2008a , p.4).

In the literature, we identified a gap that the legal aspect and public aspect are receiving less attention under the entanglement of societal, political, and technical contexts in smart cities development (Helbing & S'anchez-Vaquerizo, 2022; Helbing et al., 2021; Johnson, 2022; Buhmann et al., 2019; Brauneis & Goodman, 2018; Dubber et al., 2020; Juvenile Ehwi et al., 2022). Citizen participation have however been deemed a factor of insufficient measure in the city scale digital twin context, and explained as a mechanism that "still needs to be optimized" (Zhang et al., 2022, p.114). Habermas also shared a similar view of the political concerns when technocracy emerged during the 1960s and inspired him with the development of the democracy principle (Mingers & Walsham, 2010). The democracy principle supervises the pragmatic, ethical and moral domain and governs norms that are legally established and agreed upon by all citizens through a legislative process that conforms to legal procedures (Mingers & Walsham, 2010). The law should be able to resolve conflicts of interest in the pragmatic domain; protect goodness or virtue for the community and only include those norms that can be universally accepted (Mingers & Walsham, 2010). Each domain should carry out discussion with different stakeholders who are affected (Mingers & Walsham, 2010). This kind of discussion process resembles the participatory approach suggested by Helbing et al. (2021) and White et al. (2021) with the attitude of *Citizen First* being brought to the table to enhance the city's resilience, support city planning and policy making through citizen feedback in city scale digital twins context for long term development. The public should be able to speak up for their opinions to improve the city planning decisions that affect them and establish constructive dialogue based on solving conflicts and coming up with consensus (Helbing et al., 2021). Yet in reality, legal systems have limitations and ethical ways of thinking will act as the guidance or provide insights to improve it (Helbing et al., 2021). Dubber et al. (2020) expressed the view that ethics is the preliminary or potential form of guidance for future policy which should preferable be self-regulated due to its flexibility, instead of imposing it directly through law. That being said, the aforementioned domains make individuals aware of the socio-technical tension we face nowadays and seek solutions under various ethical viewpoints in a deliberative way (Helbing et al., 2021). The discourse can cultivate social norms or lead to the development of self-defined customs that can be applied to different situations with different cultural settings, as well as strengthening democracy (Helbing et al., 2021).
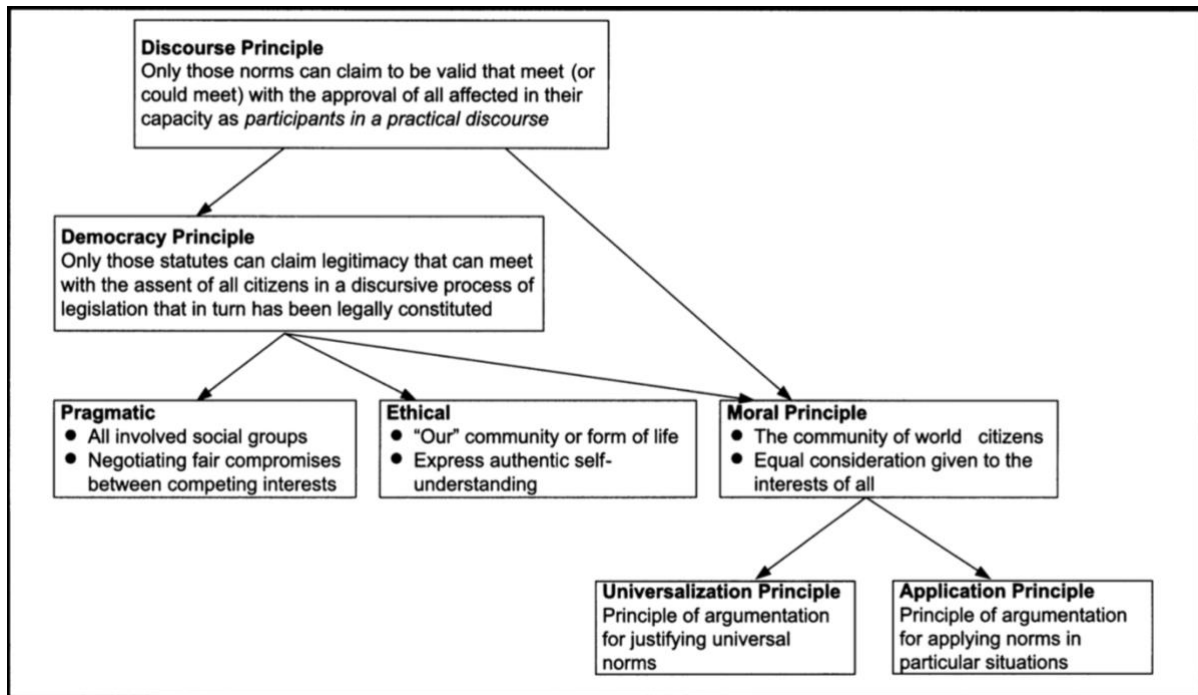
Figure 2.1: Relationships within the Principle of Discourse (Mingers & Walsham, 2010)

Underpinning the principle of discourse (Figure 2.1) that epitomized Habermas's philosophical thinking with the democracy principle, including the pragmatic, ethical, and moral domains as well as the theory of communicative action will act as a multi lens to investigate the research question of interest. Classical ethics principles can support meticulous and transparent examination of the consequences connected to the rationale behind human decisions and guide practitioners in dealing with ethical conundrums that arise from different situations (Gal et al., 2022). Hence, we emphasize the power of discourse to promote the engagement between stakeholders in order to analyse different human decisions grounded with various ethics considerations. This is also to outline the current stage of city scale digital twins in data management processes as well as the considerations when encountering algorithmic bias to provide an avenue for future discussion and improvement.

2.5 Conceptual Foundation

A conceptual foundation of applicability in relation to the topic at hand and the prevailing research question have been derived from the literature review. As the research question of interest is: "*How are pragmatic, ethical, and moral considerations prioritized by digital twin practitioners during data management processes, including collection, usage, storage, sharing and quality as well as when encountering algorithmic bias in those processes?*", the themes of most consequence for examining those ethical implications in relation to managing data in a city scale digital twin and additionally when encountering algorithmic bias have been identified and summarized as *Data Management*, *Algorithmic Bias* and *The Principle of Discourse*. Data includes the factors data collection, storage and use, sharing and quality as they were the most prominent in existing literature. Bias have been explained to consist of biased people, biased data, and biased algorithms. However, the aspect of biased algorithms will be outside of the scope of this thesis considering the authors lack of possible insight and understanding of the extensive algorithms and machine learning models used in city scale digital twins. Hence, data and algorithmic bias are the identified phenomenon of most interest in relation to city scale digital twins. The principle of discourse explains the ethical aspects contributing to a democratic endeavour with a possibility of bias mitigation. The themes are further summarized under section 2.5.1-2.5.3 below.

2.5.1 Data Management

> Data management represents the first and core theme of the city digital twin.
> (Shahat et al., 2021, p.9)

We discern the city scale digital twin design and implementation process with two distinct features found in this technological approach, which is algorithms and data. Since data-driven urbanism is the key driver in city scale digital twins, the focus of this research will be on examining the *data collection, storage, usage, sharing* and *quality* throughout the process in this research. In regard to data-driven urbanism and those aforementioned data management aspects, concerns of confidentiality, discrimination, liberty and unjust treatment are prominent (Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Brauneis & Goodman, 2018; Kitchin, 2016; Juvenile Ehwi et al., 2022; Charitonidou, 2022; Sunstein, 2022). Data in a city scale digital twin is gathered in full scale and in real-time through the employment of IoT and ubiquitous sensing, contributing to higher levels of data granularity (Kitchin, 2016) and subsequently further complexity. Data can provide a full picture of certain situations, but it may miss other societal aspects and human values such as unity, shared values and norms that are hard to measure or comprehend (Helbing & S'anchez-Vaquerizo, 2022). Furthermore, Ferré-Bigorra et al. (2022) stated that data quality is problematic as it is unevenly distributed, skewed and risks being mistaken by humans. Seeing as data is a prominent aspect of city scale digital twins that pose several concerns, particularly the risk of bias (Zou & Schiebinger, 2018; Rai et al., 2019; Stinson, 2022; Akter et al., 2021), it is important to include as a theme in this empirical endeavour to deliberatively investigate the data management process. Table 2.1 summarizes the most relevant literature found in regard to data management processes in relation to the topic of interest.

Table 2.1: The data management aspects included, and important literature connected to each factor.

Theme	Factor	Literature
Data Management	Data Collection	Goodman (2020); Hamalainen (2021); Helbing et al. (2022); Helbing et al. (2021); Brauneis and Goodman (2018); Kitchin (2016); Ehwi (2022); Charionidou (2022); Sunstein (2022); Petrova-Antonova and Ilieva (2021)
	Data Storage & Use	Goodman (2020); Hamalainen (2021); Helbing et al (2022); Helbing et al (2021); Brauneis and Goodman (2018); Kitchin (2016); Ehwi (2022); Charionidou (2022); Sunstein (2022); Petrova-Antonova and Ilieva (2021)
	Data Sharing	Goodman (2020); Hamalainen (2021); Helbing et al (2022); Helbing et al (2021); Brauneis and Goodman (2018); Kitchin (2016); Ehwi (2022); Charionidou (2022); Sunstein (2022); Petrova-Antonova and Ilieva (2021)
	Data Quality	Ferre-Bigorra et al. (2022); Papyshev and Yarime (2021); Brauneis and Goodman (2018); White et al. (2021); Deren et al. (2021); Kitchin (2016); Helbing (2021); Charionidou (2022); Sunstein (2022); Petrova-Antonova and Ilieva (2021); Juvenile Ehwi et al. (2022); Dubber et al. (2020); Helbing and S'anchez-Vaquerizo (2022)

2.5.2 Algorithmic Bias

As algorithmic bias and its socio-technical consequences is underexplored in digital twin literature (Tzachor et al., 2022) and hence the most central factor of the research gap we are aiming to address, it is important to include algorithmic bias as a theme in this empirical endeavour. Kordzadeh and Ghasemaghaei (2022, p.395) define algorithmic bias as a "systematic deviation from equality that emerges in the outputs of an algorithm". Algorithmic bias has three main drivers starting with the people, reflecting in the data, reinforcing in the algorithms (Akter et al., 2021; Stinson, 2022). Scholars are alarming the risk of societies' existing biases being amplified if those biases permeate training data and the algorithms themselves (Rai et al., 2019). Data-driven innovations risk producing "unjust, unfair, or prejudicial" (p.1) results based on gender, race, sexual orientation and more (Akter et al., 2021) and will thus be discussed in the city scale digital twin context in this thesis, both in relation to data management processes and the ethical domains stemming from the democracy principle of Mingers and Walsham (2010). As previously mentioned, algorithms are not in the scope of this thesis, meaning the theme of algorithmic bias will only include the factors, people, and data. People collect data, embedding their own biases in the systems (Akter et al., 2021; Stinson, 2022). Stinson (2022) state a need to further include minority groups in development processes, since the negative outcomes usually stems from developers' lack of perspective and knowledge. Lack of representation and diversity can further lead to biased datasets (Stinson, 2022). As no common standard is defined, the maintenance of the data quality is hampered, like its metadata and data versioning as it is all dependent to human judgment and their decisions based on different circumstances (Papyshev & Yarime, 2021).

Algorithmic bias disproportionately affects and discriminates against already marginalized groups and people (Stinson, 2022; Kordzadeh & Ghasemaghaei, 2022). Several researchers are emphasizing stakeholder engagement and citizen participation in relation to city scale digital twins (Akter et al., 2021; Dembski et al., 2020; Zhang et al., 2022; Mohammadi & Taylor, 2018; Shahat et al., 2021). Dembski et al. (2020) emphasizes consideration of citizens' experiences, including marginalized groups. Zhang et al. (2022) call for a "people-oriented and

more inclusive" (p.114) approach to city scale digital twin development. Akter et al. (2021) underlines citizen feedback and stakeholder engagement, to produce regular insight about datasets to create fairer systems that better meets the needs of the affected people. Similarly, Helbing et al. (2021) and White et al. (2021) advocate for a participatory approach, proposing a *Citizen First* mindset entering the discussion to enhance city's resilience, supporting city planning and policy making. From the democracy principle, Habermas emphasized the Theory of Communicative Action as a method to achieve consensus through communication in which everyone has the equal right to participate in the debate and nobody can exert power over others' right to make an utterance (Mingers & Walsham, 2010). The level of engagement between stakeholders and citizens will be analysed under the democracy principle and more specifically TCA in order to understand different human decisions grounded with various ethical considerations. Table 2.2 summarizes the most relevant literature found in regard to algorithmic bias in relation to the topic of interest.

Table 2.2: The algorithmic bias factors applied in this thesis, and important literature respectively.

Theme	Factor	Literature
Algorithmic Bias	Data	Helbing et al. (2022); Afrashtech et al. (2020); Gal et al. (2022); Ananny and Crawford (2018); Brauneis and Goodman (2018); Kitchin (2016); Charionidou (2022); Sunstein (2022);
	People	Brauneis and Goodman (2018); Kitchin (2016); Sunstein (2022); Dembski et al. (2020); Zhang et al. (2022); Helbing et al. (2021); White et al. (2021); Mingers and Walsham (2010); Mohammadi and Taylor (2018); Akter et al. (2021); Nocha et al. (2021); Shahat et al. (2021);

2.5.3 Principle of Discourse

According to Mingers and Walsham (2010), when achieving democracy, it is required to balance pragmatic, ethical and moral domains respectively with deliberative discourse, using discussion as engagement to reach consensus, when encountering the latest technologies and enhancing understanding of the rationale behind stakeholders' intention. Echoing to the socio-technical perspective in information systems, we believe each decision making and process are the "assemblage of human and non-human actor[s]" in creating IS artifacts (Ananny & Crawford, 2018, p.974). Thus, those decisions will be under the scrutiny of the ethical lens, more specifically the Principle of Discourse, to achieve the research purpose. Table 2.3 summarizes several common ethical principles that have been used in the information systems area and some are specifically for digital twins, smart cities, algorithmic or data-driven decisions. The Principle of Discourse consists of ethical principles being used in other areas and is considered to be a comprehensive ethical principle including pragmatic, ethical and moral domains under the democracy principle, including the theory of communicative action (TCA). The findings of the research will be discussed in relation to those three domains under the democracy principle to determine the current state of data management processes and algorithmic bias in the city scale digital twins context. It is to identify various factors that practitioners and city planners need to consider when managing data and addressing algorithmic bias during the data management processes as those considerations pose a potential trade-off of the domains.

Table 2.3: The principle of discourse applied in this thesis, and important literature connected to the domains.

Principle of Discourse		Ethical Principles	Literature
Democracy Principle	Pragmatic Domain		Ess and Thorseth (2008); Goodman (2020); Minger and Walsham (2010); Helbing et al. (2021)
	Ethical Domain	Computer Ethics; Information Ethics; Discourse Ethics; Virtue Ethics;	Floridi (1999); Ess and Thorseth (2008); Stahl (2008); Ess (2008); Goodman (2020); Gal et al. (2022); Minger and Walsham (2010); Helbing et al. (2022); Ananny (2015); Kitchin (2016); Ehwi et al. (2022); Helbing et al. (2021); Buhmann et al. (2020); Haubler (2021)
	Moral Domain	Information Ethics; Discourse Ethics; Consequentialism/ Utilitarian Ethics; Deontology; Virtue Ethics	Floridi (1999); Stahl (2008); Ess (2008); Ess and Thorseth (2008); Afrashteh et al. (2020); Goodman (2020); Gal et al. (2022); Minger and Walsham (2010); Helbing et al. (2022); Ehwi et al. (2022)
	The Theory of Communicative Action	Discourse Ethics	Stahl (2008); Ess (2008); Ess and Thorseth (2008); Afrashteh et al. (2020); Minger and Walsham (2010); Ehwi et al. (2022); Helbing et al. (2021); Buhmann et al. (2020)

3 Methodology

3.1 Research Strategy

3.1.1 Research Philosophy

According to Saunders, Lewis and Thornhill (2016), research philosophy helps us to develop knowledge in order to address issues or problems that we believe are important (p.124). The authors pointed out that the way of developing knowledge is based on our assumption and beliefs (p.124). Ugwudike (2022) suggested that social practices infuse human norms, culture and preferences into algorithms; algorithmic calculation is thereby influenced by our social practices. This notion resonated with the similar idea of the constitutive entanglement between social and materiality (Orlikowski, 2016) and recalled the idea suggested in Mingers (2015) about the discourse ethics of how one's norm or culture is shaping the reality in both the social and material world. The existence of algorithm bias cannot be easily solved as it is rooted in our daily digital life and influencing our daily activities (Richardson, 2021). However, what if algorithmic bias happens in a country or region that pursues equality like Sweden (Sweden.se, n.d.)? Will this value transfer to the city scale digital twin design and implementation process?

Our research is focused particularly in terms of understanding the rationale behind decisions and considerations in data management processes during the planning and design of city scale digital twins and how it shaped the incident. According to Mingers (2004), critical realism (CR) "wants to get beneath the surface to understand and explain why things are as they are, to hypothesise the structures and mechanisms that shape observable events" (p.100). Carlsson (2003) described CR "as a specific form of realism. Its manifesto is to recognize the reality of the natural order and the events and discourses of the social world" (p.5). CR consists of three domains which include the *Real*, the *Actual* and the *Empirical* domain (Carlsson, 2003). The Real domain is the underlying mechanism and structures that exert its influence on a social setting that enables or constraints actions; actions occur in the Actual domain with the people creating events in the domain or regenerating structures back to the Real Domain while experiences of the action lies on the Empirical Domain (Carlsson, 2003). We believe different decisions drawing from different events and stakeholders in the social setting can be comprehended through classical ethical theories and discourse that echo to CR, accepting that multi existence of realities with sharing features could be found. Therefore, we grounded our epistemological assumption in critical realism.

3.1.2 Research Approach

In order to achieve the aim of critical realism, "interpretive research takes the respondent seriously and attempts to develop a coherent account of actions that allows the audience to understand the situation in question" (Stahl, 2008a, p.8). Mingers (2004) mentioned that CR does not commit to a single research approach where the consideration should lie on whether it fits into the research aim and use. Patton (2014) suggested using qualitative inquiry with in-depth interviews can better understand the circumstances of the phenomenon (p.209). In contrast to qualitative inquiry, the quantitative methods commonly in use are surveys and experiments where control is applied by researchers through designated questions as well as answers (Recker, 2013, p.36). In addition to that, Recker (2013) stated that when the research

focus is on data or numbers, there is further a risk of hidden bias in data (Patton, 2014). In our research, we tried to identify any algorithm bias that is already introduced in the data management process when planning and designing the city scale digital twins. Quantitative methods might however not realize our aim to understand the rationale behind the actions taken by interviewees in the processes as well as the chances of encountering unforeseen bias in data and numbers through the selected questions or answers. Therefore, we proceeded with qualitative inquiry so as to extract extensive information from interviewees. Instead of aiming for controllability, we prioritized our value to obtain detailed information by using qualitative inquiry for conducting our research in order to better comprehend the activities throughout the processes to study the phenomenon in depth.

For the interview format, an in-depth interview is recommended as it can provide the maximum flexibility regarding the interviewees' responsiveness to different situations and their individual differences (Patton, 2014, p.865). Yet it "may require a greater amount of time to collect systematic information because it may take several conversations with different people before a similar set of questions has been posed to each participant in the setting" (Patton, 2014, p.865). In this regard, we opted for semi-structured interviews instead, due to the time and resources limitations. During the interview, we explicitly asked about the knowledge of algorithm bias to the interviewees and documented how they defined this topic and how it affected the decisions during the data management processes in city scale digital twins.

After we collected the qualitative data from the interviews, we used an interpretive approach to interpret the data as Recker (2013) stated that it helps to study a research topic in context through "subjective interpretation within socio-historical context" (p.88). In view of the subjective interpretation of data, we elaborated more on the scientific credibility and ethical consideration in section 3.6 and 3.7 respectively.

3.2 Literature Review

A comprehensive literature was produced firstly for the authors to retain a thorough understanding of the topic of interest, and secondly to provide a common understanding of the relevant topics for anyone interested in reading about this research endeavour. While conducting the literature review, several measures were taken to ensure an adequate review. As further discussed in section 3.6, literature included in the review should entail a certain quality of their own (Oates, 2006). To find literature of high quality, Oates (2006) proposes searching in journals with peer reviewing standards, the purpose of peer review is to assess suitability and quality before publication. To produce a reliable literature review and ensure scientific quality, journals of selected literature were scrutinized to confirm peer review standards, and heavy arguments were built on scientific articles with origin in established IS journals, for example MIS Quarterly from the IS basket of 8. To search and filter among literature, Scopus database and LUBsearch were mostly utilized which contributed to easy assessment of journal quality, relevance, and other interesting metrics.

After seeing indicators of a gap in certain literature introduced to us in previous courses on the master level, initial literature searches included search words, in combination or freestanding, such as "Digital twin", "Smart city" and "Algorithmic bias". Related literature and eventually a suitable framework was identified using a snowball method stemming from articles relevant to those initial search words. Further searches were based on newly identified concepts of

relevance, such as "Data driven urbanism", "Information systems ethics" and "Principle of discourse" in an iterative process throughout the literature review.

After locating the literature, we reviewed the existing literature and identified a knowledge gap that confirmed our initial research interest. Then we redefined our research focus in order to problematize it as our research question. Subsequently, the literature review was used to develop the key themes "Data Management Processes", "Algorithmic Bias" and "Principle of Discourse" together with the derived sub themes in our conceptual foundation.

3.3 Data Collection

3.3.1 Research Instrument

As previously stated, this research entailed semi-structured interviews. This approach allowed for enough structure to be feasible under prevailing constraints, time and resources, yet still be flexible to go deep into any theme or issue, expected or not, during the interview, as described by Oates (2006). The construction of an interview guide aided in making sure no important topics were forgotten or missed (Patton, 2014). Furthermore, it served as grounds for evaluation and testing prior to the actual interview, improving the efficiency and effectiveness of the interview (Patton, 2014; Oates, 2006). According to Oates (2006), the interview, and the interview guide in preparation, should optimally begin with easier questions for the participant to answer, to build up trust and establish credibility. Topics of greater sensitivity and difficulty should be presented later in the interview (Oates, 2006).

In our interview guide, we designed the sequence of questions carefully in order to develop the trust with the interviewees. As our research topic is surrounding algorithmic bias in the data management processes of city scale digital twins, interviewees might find this topic sensitive and be reluctant to provide information about it. Thus, starting questions included personal questions about the participant's role and work with digital twins that is easy to elaborate on, questions about algorithmic bias and its consequences were posed further into the interview. Additionally, it is preferable to use open-ended questions, starting with "what", "how" and "why", to trigger more profound answers (Oates, 2006; Patton, 2014). We echoed to this notion and designed most of our questions as open-ended questions so that the participants could elaborate information surrounding the topic, which was realized as much as possible with the exception of clarifying statements during the interviews. Those open-ended questions enhanced the richness of information and promoted some unexpected answers which led us to come up with further relevant questions related to the topic in a conversational manner. Unfortunately, this also meant that not all participants got a chance to answer certain questions specifically, which remains a limitation of the empirical data collection process. Apart from the questions, preference of languages, either Swedish or English was asked before the interview in order to let interviewees feel comfortable participating in the interview. The flow of the interviews became increasingly smoother and provided us with extensive data for analysis.

The interview should be conducted in a setting that is comfortable for the participant (Oates, 2006), to build trust with interviewees that can potentially provide higher quality data. Proposedly in the participants office or the like. Each participant of our empirical data collection participated either from their home or office, whether online or in person. To prepare for the interviews, it is recommended to research basic information about the participants and their

context, workplace as well as company reports, for the possibility of highlighting potential issues of interest as well as the ability to assess the accuracy of given answers (Oates, 2006). These measures were realized iteratively, research about the participants were automatically conducted while searching for potential interviewees, and additionally in a reviving manner before each interview. Company websites, articles or news related to the city scale digital twin projects were researched before the interview in order to understand the interviewees better which provided us the context of their work in this area. To establish credibility and professionalism, all equipment should be preparatory checked (Oates, 2006). We used the video conferencing software Zoom to conduct the interviews online. Software, computers and internet connection were checked before the interviews as well as complementing the recording through both researchers' mobile phones. When the interview was conducted in person, both researchers' computer and mobile phones were used to record the interview. Besides, the researchers also jotted notes during the interview to indicate important ideas for follow-up questions as well as specific reaction like long pause in the conversation.

3.3.2 Interview Participants

When trying to achieve in-depth qualitative inquiry, Patton (2014) argues that a small number of cases, if information-rich, can be of value. This in contrast to searching for breadth regarding gathered information, thus aiming for a bigger participant sample. According to Patton (2014), "[s]ampling to the point of redundancy is an ideal" (p.474), however, this is more commonly done when the research has limitless time and resource constraints, which is not the case in this research. Although there is no expectation to be able to generalize conclusions over a larger population, one can learn much from small samples of information-rich cases (Patton, 2014). Considering the compromise of semi-structured interviews rather than proper in-depth interviews, it might be beneficial to aim at a slightly higher sample size than what is reasonably expected from in-depth inquiry. Patton (2014) recommends stating an expected minimum number of participants, where the final number of participants usually range from 4 to 10 participants for in-depth interviews. Consequently, this thesis was aimed to include approximately 8 participants.

Deep understanding of a phenomenon typically requires purposefully selected participants that suit the endeavour at hand (Patton, 2014). It is proposed to first establish criteria to ensure interviewees meet the purpose of the research, second identify willing, criteria-meeting participants within a *program*, and lastly randomly select a reasonable number of participants from the identified pool of participants (Patton, 2014). As the implementation of city scale digital twins is still at an early stage, the *program* of the thesis at hand consisted of companies that work with digital twins towards at least one of the smart city goals identified. When identifying potential participants for this thesis, digital twin projects in the city setting were identified, and people involved in said projects were contacted and offered to participate in the research as a form of purposeful sampling. We sent 15 interview requests in total for those projects that fulfilled the aforementioned requirement within Sweden. From explaining the topic of interest briefly, some contacted people referred further to colleagues that were believed to better fit the research in a form of snowball sampling method. Finally, 9 participants were interviewed during 8 separate occasions. Participants from both a technical perspective and a city planning perspective were included for the sake of triangulation. An anonymized summary of the participants and respective interview format is displayed in Table 3.1 below.

Table 3.1: Interview participants.

Interview	Participant	Perspective	Date	Format	Duration
I1	P1	Technical	12/4/23	Digital	1h17min
I2	P2	Technical	12/4/23	Digital	53min
I3	P3	City planner	17/4/23	In person	1h14min
I4	P4	Technical	17/4/23	Digital	1h18min
I5	P5	City planner	13/4/23	Digital	50min
I6	P6	Technical	26/4/23	In person	1h37min
	P7				
I7	P8	Technical	28/4/23	Digital	30min
I8	P9	City planner	2/5/23	Digital	47min

3.4 Interview Guide

An interview guide was drafted as Patton (2014) recommended, to aid in making sure no important topics or themes were forgotten or missed during the interviews. All questions were developed based on the literature review, either questions already defined in the literature were reconfigured and used to fit this research endeavour, or the knowledge gathered while conducting the literature review was used to produce questions relevant to the topic and themes. The interview guide served as grounds for evaluation and testing prior to the actual interview, as mentioned earlier to improve the efficiency and effectiveness of the interview (Patton, 2014; Oates, 2006). This was mainly done with the help of our supervisor, to make sure the questions were formulated understandably and that the flow of questions was adequate. Also, we asked three colleagues to review the interview questions as a third-party point of view to see if they could understand the questions clearly. As the interviews were semi-structured, not all questions needed to be posed ultimately, and some additional questions were asked to clarify answers. Additionally, not all questions were posed as they were designed in the interview guide but were rather revised in real-time during the interview to fit the atmosphere and flow of conversation. The thematic overview in table 3.2 shows the structure of the themes and their interrelations, as well as which questions represent those themes in the interview guide, which can be found in Appendix 2.

Table 3.2: Thematic overview with the aggregation of the representative interview questions.

Theme	Factor	Represented in Question
Data Management	Data Collection	15, 16, 17, 20, 22, 23
	Data Storage & Use	21, 25, 26, 27, 28
	Data Sharing	24, 25, 27
	Data Quality	20, 28, 29
Algorithmic Bias	Data	5, 18, 19, 20, 21, 28, 29
	People	5, 10, 18, 19, 20, 21, 28, 29
Principle of Discourse - Democracy Principle	Pragmatic Domain	4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 21, 23, 24, 25, 26, 27, 28, 29
	Ethical Domain	4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 21, 23, 24, 25, 26, 27, 28, 29
	Moral Domain	4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 20, 21, 23, 24, 25, 26, 27, 28, 29
	The Theory of Communicative Action	3, 6, 7, 8, 13, 14, 19, 20, 21, 24, 25, 28
Background and Context		1, 2, 4, 5, 6, 7, 8, 30, 31, 32

3.5 Data Analysis

After we collected the data from the semi-structured interviews, we analysed the data. Patton (2014) and Recker (2013) suggested using software to assist data analysis (p.1049, p.93) while researchers still need to rely on their own skills to analyse the data. We used Google Pinpoint and Transkriptor to transcribe the recordings including both a timestamp and a speaker. Patton (2014) also mentioned the fact that difficulties of analysing qualitative data lies in the process to derive the meaning from the huge amount of unstructured data collected (p.1033). Hence, we referred to certain analysis methods or techniques for guidance.

In reference to Recker (2013), different analysis techniques have their strengths and weaknesses where researchers should consider the most suitable approach according to their philosophical stance. The philosophical stance of this research provides flexibility to choose the analysis approach or techniques, such as inductive, deductive or both approaches as long as it is aligned to our epistemological stance (Carlsson, 2003). As the literature review concluded in three themes consisting of 10 factors that are deemed important for answering our research question, these themes and factors were utilized as a priori, assisting us in understanding the prioritization of the democracy domains in relation to algorithmic bias in city scale digital twin data

management processes as well as the activities in the processes. Therefore, we opted for a deductive approach and thematic analysis that aimed "to discover underlying structures that generate particular events and patterns" (Carlsson, 2003, p.5). The interview transcripts were all separately coded by both authors, in the overarching themes using the colour codes, underlining important ideas, and further specified using the text codes, including every code interpreted in that specific part of the transcript. The text codes for Data Management and Algorithmic Bias were posed in one parallel column of the transcript as a first layer, and the text codes for Principle of Discourse in another as a second layer, for the sake of readability of the codes in the transcripts. An illustrating example of this can be found in Table 3.4. Although the text codes of the Principle of Discourse were specified throughout the transcripts, those interpretations were not presented in the findings, as we wanted to give an objective illustration of the activities described. Those interpretations were instead discussed as implications in the Analysis & Discussion section as a layer on top of the activities identified in the first layer. In case of colour codes overlapping, the most prominent one will be coded, supplemented with text codes. An additional code was added to identify data explaining the context of the digital twin projects under investigation. The codes used are depicted in table 3.3 below.

Table 3.3: Colours and codes for each theme and factor respectively, deductively.

Theme	Colour Codes	Factor	Text Codes
Context	Yellow	/	/
Data Management	Blue	Data Collection	DC
		Data Storage & Use	DSU
		Data Sharing	DS
		Data Quality	DQ
Algorithmic Bias	Red	Data	BD
		People	BP
Principle of Discourse - Democracy Principle	/	Pragmatic Domain	PD
		Ethical Domain	ED
		Moral Domain	MD
		The Theory of Communicative Action	TCA

To reduce the risk of misunderstanding the participants' words, the recording was used in parallel to coding the transcripts, as written conversations can sometimes be difficult to interpret. The authors later compared notes in an attempt to reduce individual biases and ensure a common understanding of the findings and interpretations respectively. When encountering differing interpretations of the data, both authors' views were explained and for the most part this resulted in a common understanding and consensus. For the exceptions when a discussion did not result in consensus, data were either put aside or explained under more than one theme or factor. After extracting and structuring codes and quotes under the overarching themes, further divided into the factors, categories were formed under the aforementioned structure,

illustrating the most important considerations from the participants. For example, under the theme *Data Management* and the factor *Data Sharing, Open Data* was a prominent category that was identified in the findings. These categories were used to structure the findings further and aid readability.

We understand that the analysis of qualitative data is based on the researchers' skill and their interpretation which may not reach the rigor level of quantitative data (Recker, 2013, p.94). In this regard, we applied triangulation in our data analysis both through maintaining dependability and generating a consistent view through discussion (Recker, 2013, p.94), and by obtaining different official information from the participants or project websites in case of confusions for comparison with the empirical data to confirm the data and uphold a credibility level which is considered as "a chain of evidence" (Recker, 2013, p.94). We believe both measures can help us to maintain the rigor of our research result. Table 3.4 depicts an example of what the coding could look like. There are ultimately two versions of every transcript that were discussed and synthesized between the authors to a final version that can be found in the appendix. The participant representing each SPK_N will be specified in all appendices respectively. The timestamp and participant number will be utilized in the findings section and during discussion to reference a specific interview like so: [P8 23:12].

Table 3.4: An example of the coding from participant P8.

00:23:12 SPK_2		
The <u>pressure increases and there's a need to do more with less, not least because climate changes.</u> So our premise, our hypothesis, is it. <u>The best way to do that is to get a better understanding of what's happening in the city, granular in in in your knowledge about what how the city is evolving.</u> <u>And based on that greater insight, you can make better decisions to save the time, the money, and make the place safe.</u>	DC	PD, MD PD, ED

3.6 Scientific Quality

The quality of this research will for the most part depend on credibility, validity and reliability of the empirical data and the study in general, as generalizability is not the most relevant focus for an in-depth qualitative inquiry (Patton, 2014). However, generalizability is somewhat aided from including different types of stakeholders, such as city planners and technical people that have different areas of expertise, as well as including several different relevant projects.

> The validity, meaningfulness, and insights generated from qualitative inquiry have more to do with the information richness of the cases selected and the observational/analytical capabilities of the researcher than with sample size. (Patton, 2014, p.473)

Credibility is dependent on the thoroughness of the inquiry, including both fieldwork and analysis, as well as the reader's view of the authors and the methodological stance taken (Patton, 2014). As previously argued for, the methodological approach was considered the most appropriate regarding the topic of study which sequentially contributes to the level of quality. Regarding the authors' credibility as inquirers, track record, experience, and status (Patton, 2014, p.942) are all aspects that we have little to no control over. However, training and

presentation of self (Patton, 2014, p.942) was positively affected through extensive preparations and transparency which was realized to the best of our ability during the data acquisition, for instance through preparation of an interview guide, research preceding every interview, and transparency regarding our novice status as researchers. Further, Patton (2014) described a number of contributing activities regarding thoroughness of inquiry, this included proper analysis consisting of both positive and negative evidence for arguments, which was realized by trying to keep a critical argumentation throughout the thesis and connected to every theme and literature.

Staying focused on the research topic at hand is essential to effectively gather data of high quality (Patton, 2014). This also applies to the literature selection and review. In addition to selecting literature that is appropriate for the research topic, the literature should also retain a certain level of quality on their own (Oates, 2006). To find literature of high quality, Oates (2006) proposed searching in journals with peer reviewing standards, the purpose of peer review is to assess suitability and quality before publication. Peer review was considered throughout the entire research endeavour. Attention to relevancy and quality of the source of the literature helps to improve the quality of the study overall (Oates, 2006; Patton, 2014). Above qualitative literature, accurate referencing has been of highest priority throughout the research process (Oates, 2006; Patton, 2014), which in this research was aided using a reference management system named EndNote 20.

3.6.1 Validity

Validity is assessed based on how well the study ultimately pictures the phenomena examined in a fair and precise manner (Patton, 2014). To "picture the empirical social world as it actually exists to those under investigation, rather than as the researcher imagines it to be" (Filstead, 1970, p.4 cited in Patton, 2014, p.115). Differences in training, values, and biases in general and between the authors was sorted out and explained in order to conduct a transparent, systematic and truthful qualitative inquiry in accordance with Patton (2014), which was addressed as a continuous discourse between the authors throughout the entire research endeavour. The extensive literature review was another attempt to ensure validity, as it provided a deeper understanding of the subject of interest, reducing the risk of describing a concept in an untrue manner, establishing a common view of the subject at hand, and additionally letting the reader see our point of view. It was further represented in the empirical data findings, and in the analysis of the empirical data, contrasting views back and forth, both between the authors and compared to the literature review, to finalize a unified view of the findings.

Another notion that built our chain of evidence, as stated by Recker (2013), was to base our empirical data collection on concepts from the literature review. Furthermore, the findings were thoroughly referencing the transcripts from each interview respectively, to sustain a transparent and impartial recollection of the empirical data. The chain of evidence was utilized to "minimize bias, maximize accuracy, and report impartially" (Patton, 2014, p.179). In addition to being transparent during all phases of writing the report to showcase any deviation of rigor, triangulation has been used to maximize rigor (Patton, 2014; Oates, 2006), entailing the use of sources such as company websites and project description to verify and triangulate any collected data from individuals (Oates, 2006). However, those documents cannot be included in the research as they would contradict confidentiality in terms of participants' right to anonymity, representing a slight deviation of validity.

3.6.2 Reliability

Reliability is assessed based on the extent to which the research endeavour could be recreated by other authors and receive similar results (Oates, 2006). To aid reliability we tried to consistently explain how we conducted this thesis, upholding transparency throughout all chapters. On that notion we included the coded transcripts as appendices, we additionally included thorough references to the transcripts in the findings, for possible examination by the readers. Overall aiming at presenting a clear description of every step of the research endeavour.

3.7 Ethics

When considering ethics in this research, Patton's "Ethical Issues Checklist" (2014, pp.725-728) was consulted, holistically including 12 points. Ethical reflection and transparency (point 11) is ultimately aimed at with this chapter. Only the relevant points will be discussed, including explaining purpose (1), keeping promises to participants (3), confidentiality (5), informed consent (6), data access and ownership (7), as well as data collection boundaries (10). The remaining points: reciprocity (2), risk assessment (4), interviewer mental health (8), ethical advice (9) and ethical versus legal (12), are believed to be of less relevance for this thesis considering our position as students and the limited resources that follows (point 2), as well as the research topic at hand not requiring the level of sensitivity that points 4, 8, 9 and 12 immediately calls for.

An overarching concept to consider is the rights of participants, it includes multiple points in Patton's (2014) checklist: informed consent, explaining purpose, confidentiality, data access and ownership among others (Oates, 2006). After confirming the voluntary participation in the interview from the participants, a consent form was distributed including the participant's rights along with a description of crucial information about the thesis in the participation invitation, to ensure the participants' ability to properly give informed consent. Additionally, before we started recording the interview, we asked our participants for their consent to record the interview again to ensure their rights were being addressed. The aim was for this to be clearly understandable to the participants and give them a feeling of control over their participation (Oates, 2006), to ultimately eliminate any feelings of forced participation and positively conduct an ethical investigation. To further ensure voluntary participation, the participants' right to drop out of the study at any moment, together with their data (Patton, 2014; Recker, 2013; Oates, 2006) was clearly communicated to the participants. The consent form that has been used during this research inquiry can be found under Appendix 1.

The aspect of confidentiality is another right of the participants' (Oates, 2006). To account for confidentiality, clear boundaries were communicated to the participants about how, where and when data and general information about participants was to be stored and used (or not). In conjunction, it was in our best intention to be respectful towards participants and not go to unreasonable lengths to gain access to data (Patton, 2014). The aim was to keep a positive relationship with the participants and avoid discomfort, and rather leverage the trust built to gain access to data in a respectful manner. Additionally, the participants have the right to be narrated in an accurate and fair way (Oates, 2006), which is why we recorded the interviews, and separately interpreted the results. To maintain professionalism and the trust of participants, any intentions or promises were kept to the best of our abilities, in accordance with point 3 in Patton's (2014) ethical issues checklist. For example, sending information about the questions with good notice before an interview when that was promised.

Beyond participants' rights, it is important to account for ethical responsibility when writing the report (Recker, 2013). To avoid plagiarism, and ensure no author or researcher referenced in the report is displayed inaccurately or their work not properly credited (Recker, 2013; Patton, 2014). The report has not used others' work in an unfair way or been unrightfully crediting, to the full extent of our individual abilities. To aid this, both authors have read and discussed all the material used for the thesis.

4 Findings

The empirical finding includes 8 interviews with 9 different participants. References will refer to the specific participant and under which timestamp the information can be found, for example [P1 32:53]. To facilitate readability, repetitive phrasing and connecting words such as "ehm" and "like" has been removed from interview quotes, which means they might not be searchable in the appendices.

4.1 Context

Before presenting the findings based on the themes identified in the conceptual foundation, *Data Management* and *Algorithmic Bias*, we would like to provide some important context. Seven out of eight digital twin projects that we came in contact with during the empirical data collection were proof of concepts, several participants explicitly stating so [P2 03:43; P4 01:12/05:19; P5 18:10; P6 41:39; P7 09:42; P9 01:20/05:38]. Working towards different goals and digital twin application areas, but mainly investigating how a digital twin can be used to support their goals.

Five participants saw the digital twin as a good communication tool [P1 13:17; P3 15:12; P4 15:41; P5 10:49/13:43/24:17/45:17/45:50; P7 12:42], aiming to provide common ground [P7 16:01/17:41/01:26:19/01:26:41] and understanding between stakeholders [P4 15:41]. P3 also emphasized the goal of creating synergies between stakeholders, by creating networks and stimulating cooperation in relation to the digital twin [P3 16:30 / 38:29 / 57:11]. A similar identified goal is efficiency, by simulating and trying things out before constructing it [P1 09:39], to find solutions that are smarter [P3 01:16/01:36/31:46] and different [P3 02:49/53:41]. Visualizing and conducting analysis [P2 07:53; P5 02:47; P7 03:00/12:42/18:42] to support decision making [P2 02:13; P5 04:53; P7 13:03]. Three participants also mentioned streamlining services for both private businesses and the public [P1 34:45/36:10: P6 35:01; P7 29:31], and to create an unbroken digital chain for urban planning [P2 07:53]. Mirroring P3 and the goal mentioned about being more company friendly and service minded [P3 07:28/09:10].

Several identified goals are related to sustainability [P4 17:55; P5 08:22; P8 23:12; P9 03:05], for instance predicting where to optimally place solar panels [P1 15:13] or creating more climate friendly solutions and reducing CO_2 emissions [P3 05:31/06:30/10:47/13:23/16:30]. One participant emphasized CO_2 as a new type of currency, slowly incentivizing companies to change their models and structures to become sustainable compared to their original systems, although complexity slows down the process [P3 29:55]. P8 stated urbanization as a driver of the digital twin development area and an increasing "need to do more with less" in connection to climate changes [P8 23:12].

P8 further expressed the motivation to "develop [a] city around people as opposed to around the vehicles providing spaces for people to exist" [P8 23:12]. Another goal connected to the digital twin is increasing wellbeing and liveability in the society [P1 02:59 / 09:39; P3 07:28; P4 32:41; P6 24:39; P7 25:47; P8 01:57; P9 03:05]. Aiming to make the city safer [P1 13:17; P5 08:22; P8 01:57; P9 03:05], and a nicer place to live [P6 27:33: P7 25:47; P8 01:57]. There is an ambition in three out of eight projects to use the digital twin to make city planning wiser [P1 07:08 / 09:39], to examine different solutions before implementing in the physical space

and discover what works and what does not [P3 33:50; P6 27:33; P7 25:47]. That includes increasing city resilience by planning for crisis situations [P1 10:32; P8 00:18/01:57]. "The main goal is for the citizens" [P3 14:24], to "make the life better for the citizen[s]" [P1 23:53 / 25:11]. For instance, cleaning streets, avoiding noise pollution and traffic jams, the goal is for everything to work smoothly so the citizens don't even think about it [P1 24:19 / 25:14 / 25:52]. Identified limitations in regard to the goal of well-being is the need to balance around citizens' differing perceptions of quality of life [P3 07:28], as well as the fact that well-being is hard to quantify [P7 01:24:31].

Four participants questioned the definition of a digital twin [P1 00:24 / 02:59; P2 04:23; P7 05:07; P9 01:20]. Two participants referred to a digital 3D model of a city as the most common stage of digital twin projects, without IoT and sensors directly connected to it [P1 13:17; P2 30:54], P1 stating "I don't think any city in Sweden has that yet?", further discussing the fact that everyone has different perceptions, which makes discussions related to the digital twin more difficult [P1 33:17]. Participants expressed the need to learn and build a competence around the digital twin [P2 07:12; P5 21:07; P9 02:29], and daily work processes with the digital twin is seen as something distant [P5 13:43].

It was clear from both P1, P3 and P5 that the maturity of the technology is low, mostly in the development phase, stating it's a long way to go before they could use the digital twin for traffic planning overall the city [P5 13:43] and "we're not there yet" [P1 25:51]. Slow organizational change was listed as a factor that slows down the development of digital twin projects, including the difficulty to adapt to new technologies [P1 09:04]. Reflecting to every municipality having their own governance to an extent, leading to slow adoption of common standards if it is voluntary change [P1 45:53 / 38:55; P3 15:57]. P1 also expressed a need to "find a common core" of the data that can be reused and standardized to improve efficiency and up-scaling of digital twin projects on a national level [P1 43:01]. P3 further elaborated on open-source resources, as a way to increase the development pace since it requires less money and funding, making it easier to get started [P3 38:29]. P8 described the different layers of a digital twin meaning that development is mostly in the lower levels.

> You could think of different layers in that digital twin and at the bottom you have the kind of built infrastructure, the kind of the buildings and the roads, and that changes at a certain rate. The next layer up, you have traffic signs and park benches and trees that is still almost static but changes much faster. Because trees grow and traffic signs fall over and then the next layer up you have much more dynamic data like vehicle flows and pedestrian flows and activity and things like that. And I think today we're we're at the base layer and we're moving up to the to the middle. [P8 27:11]

4.2 Data Management

4.2.1 Data Collection

4.2.1.1 Type of Data

The data obtained in the participants' different digital twin projects were mostly statistical data, specifically in the type of geographical data [P1 47:08/50:43; P2 33:13; P3 01:36/36:07; P7 05:07], traffic data [P4 31:41; P5 30:18], or infrastructure data [P8 01:57]. Although P9 stated they showcased the integration of live data in one project, none of the projects integrated real time data on a regular basis. However, P4 expressed real time data as an important feature, emphasizing the function of connectivity between the city and the digital twin [P4 32:41]. P8 also explained the collection of image data as a core function in relation to their digital twin, as "a camera is a very rich sensing modality, .. there's a lot of information in an image" [P8 22:13].

4.2.1.2 Purposefully Collecting Data

Although one participant expressed an aim to collect as much volume of data as possible [P3 25:38], several participants refer to the consideration that data collection should relate back to the aim of the project [P2 23:18; P3 03:34/05:31/06:30/01:01:45; P4 32:41/01:03:51; P6 52:50; P9 06:59], P2 stating that more data than necessary for the specific purpose will only increase processing times [P2 23:18]. Similarly, another participant connects collection of data and data structure with investigation of needs [P1 02:59], explaining that the project is focused where changes will affect the most people [P1 11:48].

> How can the data from the city be the foundation for such an analysis? What need the data to be like? How can we collect the data? How can we store it? [P1 02:59]

4.2.1.3 Aggregated Data

P7 explained the process of aggregating data in several stages, to prevent certain misunderstandings of what is reality and what is only visualizations of the future [P7 51:13], or in the collection stage, aggregating traffic flow from a period of data collection to a yearly basis [P7 36:49]. Four participants further expressed awareness about patterns in the data that might be sensitive and identifiable under certain circumstances [P3 01:02:16; P4 38:37; P5 33:09/35:18; P9 34:18]. P3 stating the solution for such an issue is a discussion from time to time between stakeholders [P3 01:04:25] while several participants explained the process of aggregating data to counteract those identifiable patterns [P3 01:03:24; P5 33:09/35:18; P8 16:55]. Contrastingly one participant stated measuring movements is not related to specific people, hence unidentifiable [P2 51:40].

P3 posed an illustrative example:

> We say that zip code would be anonymous enough, but some has their own zip code, so you know, everything delivered there is just one customer, and that could be a problem. It's not GDPR, but it still could be sensitive data, .. if they have a plant or they have their own zip code, .. the data would be identifiable. [P3 01:02:16]

Pattern recognition in data was further elaborated by P9 in regard to national safety, exemplifying that a "foreign country could understand where some installations are that are important for national security" [P9 34:18].

4.2.2   Data Storage & Usage

4.2.2.1   Managing Data Carefully

Four participants expressed concerns when handling identifiable data and people related data [P1 51:32/52:25; P2 51:54; P5 31:21; P9 09:13], stating that it has "a lot to do with the intentions of GDPR" [P9 09:13]. P8 stressed the fact that GDPR is a well-intended regulation, however, both P8 and P9 emphasized the complexity of GDPR as a reason stakeholders might be scared of collecting and managing personal data [P8 16:55; P9 20:08], as "people in general .. want to follow the rules" [P9 20:08]. Ultimately there is a need to educate stakeholders, to be informed of the risks in using people related data, and to be careful with consent [P8 16:55]. P9 elaborated on the topic, stating "it's more complicated than just GDPR, it's about your intentions" [P9 31:22], predicting discussions related to other ethical issues surrounding data will receive more attention further [P9 29:02]. Mirroring P9, P4 emphasized the importance of understanding the responsibilities regarding data in relation to integrity, and having everything in order when people are going to be affected [P4 27:01/01:06:14]. Questions arising such as "how much data should we .. collect and where should we really stay out of this? In what way should we use it?" [P9 29:02], in order to understand the limits and develop clearer directions in that regard [P9 31:22].

4.2.2.2   Nudging

Three participants elaborated on the concept of nudging towards certain choices, specifically for the purpose of sustainability [P3 15:12; P4 55:58; P9 24:06]. Two participants clarified the fact that they cannot force anyone, but rather inspire people to make more sustainable choices [P3 15:12; P4 55:58], for example by displaying more sustainable options for transportation [P4 55:58; P9 24:06]. P3 also mentioned policy as a way to nudge companies into more sustainable directions [P3 15:12]. P9 emphasized the ethical considerations related to data based nudging, "how much should we nudge" [P9 29:02], acknowledging the debate "whether and how much .. a government or local government should nudge or influence their residents" [P9 24:06]. In response to the question about the risk of limiting people's choices, P7 explained that the digital twin can produce more potential options than what they previously could by hand [P7 01:23:11], P3 also stressed the aim of the digital twin to test different scenarios than are present at the moment, trying to avoid limiting people's choices [P3 53:41].

4.2.2.3   Data Interoperability

Six participants discussed the issue of data interoperability. Data is stored in siloed domains lacking data connection and interoperability [P1 15:13 / 15:47 / 34:45; P7 40:04]. Participants express consideration when transferring data between different systems [P2 15:04; P4 01:03:51; P9 08:17], geometry-, attribute-, and storing-wise [P1 01:07:59/01:08:56]. Exporting data to another system, with a different structure, "you of course have to be very careful not to sort the apples with the pears" [P3 40:48]. Mostly it is an issue of communication and a lack of meta-data, making it hard to understand the data and transfer the meaning of the data, which might make the data difficult to apply onto new purposes representatively [P1 01:14:13/01:12:39]. P2

and P7 pointed out the need to harmonize data, especially regarding potentially using IoT and real time data [P2 50:04; P7 40:45].

4.2.3 Data Sharing

4.2.3.1 Open Data

Three participants were positive to have datasets open [P1 07:08; P3 59:27; P4 50:44], while two participants already have portions of open data [P6 35:01; P7 35:08], seeing it as a more efficient way to share information, saving employees' time [P7 01:29:45].

> [A] private company that owns data .. will keep a barrier from using the data if you need to take some sort of payment for it, and the data that should be the ground for these analysis should be collected and given out from a public side. [P1 07:08]

However, at this point, the trade-off between keeping the data open, and losing a source of income, is still under discussion [P1 18:01; P4 50:44; P7 35:08/35:47], especially since taxes is a source of funding in the municipalities [P 35:08/35:47]. P4 expressed a similar notion, the possibility to combine data together with others' data without putting weight on the potential differences in value between the datasets, as that would be more efficient and beneficial for all [P4 50:44].

P9 explained that wen data is collected for one purpose, then the parties can use the data for free according to the specific regulation, otherwise they will need to pay for retrieving the data for use [P9 15:44]. P9 further mentioned GDPR and other regulations such as the Data Act as important factors in relation to open data, emphasizing new responsibilities to manage data within organizations [P9 17:32/34:18]. Connected both to privacy, and the notion of national security, revisiting pattern recognition in relation to open data, exemplifying that a "foreign country could understand where some installations are that are important for national security" [P9 34:18]. Five participants explained that some data are procured from third parties or shared by business organizations [P2 41:32; P3 58:28; P5 46:22; P7 33:37/33:56; P9 15:44/32:27], rather than the public institution themselves owning the data [P1 07:08; P6 31:31; P7 32:00; P9 15:44]. In light of this, another contrasting factor that keeps stakeholders from obtaining data for their digital twins is the "fear [from the business side] that the data will end up somewhere where it shouldn't be and they could be misused", revealing business secrets [P3 28:35]. Companies also need to ask their customers if they can share the data with another organization or business [P3 01:10:49].

4.2.4 Data Quality

4.2.4.1 Standards & Guidelines

Regarding data quality, one participant expresses a need for a national standard, considering data management in general [P1 01:03:37], for instance how data and metadata is dealt with [P2 41:32; P3 01:11:54]. Since there are several standards in use regarding metadata that differ from each other, it is hard to trace back the metadata [P6 01:15:01; P7 01:15:31], possibly harming the data quality [P6 01:16:00]. There are further very few standards for ascertaining the data quality when the data was collected by others, which may lead to inefficient data

procurements [P9 36:18/37:01/37:59]. The specific standard in use is also affecting the level of detail (LOD) on the data, and multiple differentiated standards exist [P2 07:53/13:01/44:30]. Decisions surrounding the LOD in data are explained to be challenging, considering different aspects such as collection, quality, usage and storage [P1 30:35], making it difficult coming to a consensus regarding those aspects [P1 48:03]. P2 expressed a need to "find the least denominator that fits all?" [P2 23:18].

When it comes to the actual collection of data, the existing guidelines are mostly on how to collect data, not where to collect data, meaning the line between having very specific data and no data at all is blurry [P1 55:42], affecting the level of data quality in certain places. P5 explained the use of internal statistical data collection methods, including established requirements for the distribution of data points, mostly assessing the quality based on their internal expertise [P5 41:35]. P8 also stated the assessment of data quality takes place internally, stressing that data has to be of high quality as it affects the trustworthiness of the decision generated from it, which is an important aspect for the purpose of their digital twin in relation to their customers [P8 07:02].

Participants contrastingly saw possible confidentiality and privacy concerns with data being in too much detail, from a perspective of businesses protecting their data, referring back to the zip code example mentioned earlier [P3 01:03:24]. Which also plays into protecting people from pattern recognition [P3 01:02:16; P4 38:37; P5 33:09/35:18; P9 34:18] and national security on a greater scale [P1 52:25].

4.2.4.2   Data Accuracy

The accuracy level needed is dependent on how secure the analysis has to be, as it might differ between different scenarios [P2 26:29]. P2 gave an illustrating example talking about sloping ground and water flow:

> The more slope, the less accuracy you need, because then all the water obviously flows down the slope. But as the ground becomes flatter, as most cities are, the more detail you need to have on your data in order to know where the water is going to flow or when it's going to stop. [P2 23:18]

P8 stated they will accept accuracy levels no lower than those fulfilling the value proposition and data trustworthiness discussed with their customers [P8 11:37]. P7 stated as an example that it may be acceptable to have 5% inaccuracy on the data as "you'll get a bigger picture of the whole thing", if it's being communicated and understood by the people basing decisions on the analysis and outcome [P7 58:51]. To explain accuracy levels and its implications on the decision making, metadata is important [P6 46:57; P7 01:04:57]. Thus, there is a need to understand the data and describe the accuracy levels [P2 41:32], making sure to communicate it in a manner that is easily understood by the recipient [P6 47:31]. Ultimately, it is important to describe it well enough for the person making the decision to feel confident about that [P2 41:32]. However, P2 saw a lack of proper ways of describing accuracy levels [P2 41:32], P7 similarly brought up the lack of sufficient metadata when sharing data between stakeholders, decreasing the understandability of the data [P7 01:13:08].

Three participants mentioned money as a factor making it difficult to maintain high quality data. P2 explained that the specialists decide what accuracy levels are needed depending on the specific scenario, but ultimately the level of accuracy is dependent on the decision maker's

willingness to pay [P2 28:30]. P7 explained that the last 5% of accuracy from 95% to 100%, are the most expensive ones [P7 58:51], further stating that it is impossible to get 100% accuracy on data [P7 01:03:29].

4.2.4.3   Data Versioning

One of the participants expressed the view that it is challenging to keep the data up to date and work on the data versioning [P1 06:14] as they can only update the data when they discover issues, but the errors may have existed for several years [P1 28:29]. P2 explained it as an organizational problem, versioning in silos [P2 20:54], appointing no clear roles and responsibilities related to the data versioning update process [P1 26:28/27:36/28:04]. P2 further connected the consequences of bad versioning back to financial aspects:

> As soon as the information is not updated and not actual, people stop trusting it, the users won't trust it .. then it's throwing away your money. [P2 20:54]

P6 and P7 expressed the priority of up-to-date data rather than high accuracy data specifically for the purpose of the digital twin, it matters less to know if a data point is one centimetre wrong, what matters more is if it is still there [P6 01:06:33; P7 01:04:57]. P7 explained an illustrative example in relation to decision making. In a case of the digital twin being utilized to decide whether a bus stop is close enough according to certain requirements, for a person living in a neighbourhood on the other side of the train tracks, it is important to understand the path the person would have to take. "If you live here and the bus goes from here, it's only 80 meters away [bird way], but it's four kilometres to go there" [P7 27:42]. Giving the example of basing a decision on data that was collected before a tunnel was built underneath the train tracks, significantly shortening the path to the bus stop, affecting the sense of the decision [P7 45:02]. Metadata is continuously important to document those issues, to describe when and how the data was collected [P7 01:04:57].

4.2.4.4   Data Representation

> It's just as important that somebody reliable can handle this data and put it into the digital twin in a reliable way, in a way that mirrors reality. [P3 22:36]

To ensure high quality data, you need to understand the data, otherwise it is difficult to see what the data actually represents and how you can utilize it [P9 13:35]. P9 mentioned doing quality checks to ensure representative data, for example when collecting traffic data, they want to understand what the vehicle fleet represents, sports car vehicles and moped vehicles would represent a "completely different speed range and also the coverage of data" [P9 37:01]. P8 similarly pointed out the importance of having "evenly distributed training data, to have "enough representative examples of each class" [P8 08:18]. If there is a lack of understanding about the data the tendency identified at the first glance might not be correct, making any decisions based on that data faulty [P6 01:09:54; P7 01:07:26/01:13:08]. Contrastingly, P9 expressed the need to understand the data in relation to the purpose, sometimes the "causality is not always that exact, and it doesn't have to be, maybe it's good enough" [P9 40:48]. This ties back to the issue of metadata, to properly describe what the data is about and how it was collected, to improve solidity of the decision making [P2 41:32; P8 07:02]. P7 stressed a need to be aware of what the data represents and what implications that has on future decision making, in relation to the example of the bus station and understanding the path a person would have to take [P7 45:02]. On this notion, P4 saw the digital twin as a tool to identify missing aspects of the data [P4 07:39], as for instance pedestrian behaviour can be complex and difficult

to recreate [P4 38:37; P7 36:49] the visualization in the digital twin can aid identifying misrepresenting data [P4 07:39].

4.3 Algorithmic Bias

A majority of the participants showed some level of understanding of the concept of algorithmic bias, despite variations in knowledge of the term itself, three participants had never heard the expression before [P1 01:00:25; P5 37:53; P9 23:03]. After explaining the concept of algorithmic bias briefly, two participants expressed that algorithmic bias is not part of the discussions surrounding the digital twin yet [P1 01:01:51; P3 56:30]. "We're not in that kind of detail at all" [P1 01:01:51] or "we're just not there yet" [P3 56:30]. However, seven participants mentioned some aspect of the concept of algorithmic bias in relation to the digital twin project they were working with. When discussing algorithmic bias in the interviews, the participants mostly focused on the data aspect of algorithmic bias, connecting it to data quality [P2 45:48; P4 47:02; P5 39:57; P8 10:40], data accuracy [P2 48:53; P6 01:01:43; P7 58:51], data representation [P4 47:02; P9 12:51], or imbalanced datasets [P1 01:01:14; P3 51:09; P8 09:11]. Only one participant explicitly mentioned biased people, building existing biases into a system [P3 50:26], however, traces of people bias, negatively or positively, were implicitly illustrated during all interviews.

4.3.1 People

P2 reflected on data accuracy, saying the biggest problem is not the data in itself but what people interpret from it.

> What I've seen and heard more as problematic is that inexperienced specialists are using data that .. they are not understanding fully, .. not understanding what effect that loss of quality has on their final analysis. [P2 45:48]

P7 similarly described data accuracy as something people should be more aware of, seeing a problem of people interpreting the data as the truth when you cannot get 100% accurate data [P7 01:02:56/01:03:22/01:03:29]. P2 explained it might depend on communication or metadata but emphasized lack of experience [P2 48:08], and proposed teaching affected stakeholders about "the insecurities or inaccuracies or risk levels that you are handling in your analysis" [P2 48:53]. P4 similarly expressed the problem of people not understanding about the data they consume, especially when they did not collect the data, but contrastingly stated that people that work with information understand that they have to be careful when making consumptions for a different purpose than the data was collected for [P4 01:03:51].

4.3.1.1 A Lack of Standards & Guidelines

From the technical side, stakeholders are evaluating the data quality based on internal frameworks with the goal of providing trustworthy data to the customer [P8 07:02]. The municipality also has their own methods for collecting data of high quality [P5 41:35] and further try to standardize decision making, to be equal for everyone [P6 01:02:54; P7 01:01:46], making the analysis safer and more time efficient [P6 19:50 / 20:11 / 20:35]. Two participants described specialists as drivers when it comes to decision making, deciding what accuracy levels are needed in the analysis [P2 26:29], and taking the final decision based on previous

experience of good solutions [P7 42:35]. Contrastingly, two participants expressed a need for national standards, considering data management in general [P1 01:03:37], and metadata [P1 01:14:13/01:12:39; P3 01:11:54]. It is further emphasized that when standards are lacking, humans are left to make decisions, usually at a fast pace [P1 01:11:15].

> People take decisions. That's what we are good at. Which means that if you and I do the same job, we would come up to different decisions. Definitely that's a risk of dealing with reconstructing data. [P1 01:10:18].

4.3.1.2   A Focus on the City Centre

One participant explained that the project focus is firstly aimed at where it makes the biggest difference for the most people [P1 11:48] which is usually in a city's centre [P1 13:07]. There is "always" a discussion around where it is worth spending money on changes [P1 58:54 / 59:01], taking the example of the outskirts of a municipality not being the first priority for updating data preceding a decision [P1 57:34]. P9 described the aim of the digital twin to eventually represent the entire city [P9 10:01], but for reasons such as business opportunities and tourism, with interests both from industry and academia [P9 11:04]. As a consequence of this, P9 explained that the digital twin contains more information from the central parts, stressing that people involved in the project realize this and know that the goal is to comprehend all parts of the city [P9 10:37].

4.3.1.3   Citizen Inclusion

Several participants saw value in including citizens in the discussions surrounding the digital twin [P3 54:40], as a source of feedback [P4 07:39; P7 53:29], two participants expressed citizen engagement and society involvement as a driver of the digital twin project [P1 13:17; P4 01:02:12; P9 27:44]. Only two participants, working in the same project, explicitly mentioned managing citizens' feedback regularly in the digital twin project currently, stating "it's a democracy" [P6 55:03; P7 55:08], but most digital twin projects that were discussed in the interviews had not reached that stage of the development yet. P3 drew a parallel between citizen engagement and different stakeholders working in the digital twin project being able to relate to being just a citizen as well and contribute that perspective to the discussion [P3 54:40].

4.3.2   Data

As previously mentioned, when discussing algorithmic bias in the interviews, the participants mostly focused on the data aspect of algorithmic bias, connecting it to data quality [P2 45:48; P4 47:02; P5 39:57; P8 10:40], data accuracy [P2 48:53; P6 01:01:43; P7 58:51], data representation [P4 47:02; P9 12:51], or imbalanced datasets [P1 01:01:14; P3 51:09; P8 09:11]. P4 explained the process of discussing biased data, when they realize that they are always working with biased data and have to adapt to that [P4 47:02]. Trying to answer questions like "Is this good enough, for what purpose? How can we put metadata on this so people understand what is this kind of data and who is responsible for this data? How often is it updated, how?", with an aim to understand the complexities of data in the digital twin [P4 47:02].

> Because you can use very good data and you can use very poor data and they can all be a part of the digital twin, the tricky thing is to understand which is which.
> [P4 47:02]

P4 contrastingly stated that working with a city entails continuous conflicts of interest, and satisfying everyone is a difficult task [P4 15:41], P3 further stated that "if we will have consensus on everything we do we won't get anything done", emphasizing a trade-off [P3 01:11:54]. P4 stressed that despite those difficulties, the digital twin can aid " to see the whole picture" [P4 15:41].

Most of the findings related to biased data have been presented and explained in the *Data Management* section of the findings. The findings that are of importance in regard to biased data include the aspects of blurry lines geographically between high- and low-quality data [P1 55:42] and data quality in general, poor data versioning leading to presenting a false picture of reality [P1 06:14/26:28; P2 20:54; P6 01:06:33; P7 01:04:57], the issue of unclear metadata, making consumptions of data difficult and risky [P1 01:14:13/01:12:39; P2 41:32; P3 01:11:54/40:48; P6 01:15:01; P7 01:15:31; P8 07:02], discussions surrounding data accuracy [P2 26:29/28:30; P7 58:51/01:03:29; P8 11:37] and the trade-off of aggregating data in favour of integrity [P1 01:01:51; P3 01:03:24; P5 33:09/35:18; P8 16:55; P9 34:18], and understanding the representation of the data [P3 22:36; P6 01:09:54; P7 01:07:26/01:13:08; P8 08:18; P9 13:35/40:48], "maybe the data doesn't really represent what you thought it would represent, it represent something else" [P9 13:35].

P3 pointed out that the municipality aims to collect as much data as possible to create a fair representation of reality, but contrastingly expressing that data collection is still limited at this point in the project [P3 25:38]. P7 also explained that the last 5% of accuracy from 95% to 100%, are the most expensive ones [P7 58:51]. There is further a common understanding that money is of significant influence in the digital twin projects [P1 16:30; P3 02:49 / 30:51; P4 17:55; P8 04:17/23:12/24:57]. Both in terms of managing internal resources [P4 17:55; P6 23:31; P7 29:31; P8 04:17], to "have that greater insight, better updates and then more effectively manage their resources to respond to issues and problems" [P8 00:18], and in terms of making priorities in relation financial implications, if "something is very expensive and doesn't make the big change then you won't do it" [P1 16:30]. Ultimately the level of accuracy is dependent on the decision maker's willingness to pay [P2 28:30]. It is also stated that politicians influence the goals of the digital twin [P6 44:49; P7 09:20; P9 26:34] because they manage the taxes [P1 17:33 / 18:01]. Another financial aspect influencing the digital twin projects from the governmental side is the Public Procurement Act, influencing the municipality to choose the cheapest vendors that fulfil their basic requirements, with the risk of ending up with something that is hard to work with [P1 39:53 / 40:44 / 41:24 / 41:27].

5 Analysis & Discussion

Various socio-technical challenges involved in city scale digital twins are both complex and controversial, involving conflicts of interest, stakeholders and agendas, researchers suggested to examine it from an ethical perspective to understand the rationale behind each action in a critical and rational way (Juvenile Ehwi et al., 2022; Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Dubber et al., 2020; Buhmann et al., 2019; Häußler, 2021; Tzachor et al., 2022). Since the issues are so complex in nature, it is hard to reach a consensus regarding to what extent choices are considered as optimal or acceptable and what will be the trade-off or positive outcome. In this section, we elaborate on the activities that were identified in the findings that are considered as complex and controversial which cannot be easily interpreted.

For instance, algorithmic bias can exist spanning the data management processes (Akter et al., 2021). In the literature review, we mentioned the conundrum between cost and worthiness and how it affects data quality as well its data accuracy surrounding the processes in data collection, storage, usage and sharing (Ferré-Bigorra et al., 2022). It is integral for digital twin practitioners to be accountable for their decision and action deliberately when handling data and identify the possible outcome from those decisions taken. Therefore, we are not defining the rights and wrongs in this endeavour but rather describing how the democracy domains within the principle of discourse are represented in the discussions when planning and designing city scale digital twins. By utilizing the principle of discourse (Mingers & Walsham, 2010), we are going to examine pragmatic, ethical, and moral aspects under the democracy principle and the theory of communicative action with regard to human choices in those complex activities.

5.1 Data Management Processes

Papyshev and Yarime (2021) pointed out that there is no standardized procedure or guidance for data collection in general which may affect data quality. The authors also pointed out that multiple standards are being used for certain types of data such as geographical data. Several responses also resonated with this idea and called for establishing national standards or common guidelines for data collection being shared among practitioners [P1 01:03:37; P2 07:53/P3:01; P5 38:56/39:57; P6 01:15:01; P7 01:15:31; P8 07:02; P9 36:18/37:59]. Although GDPR is established to guide practitioners handling personal data (Intersoft Consulting Services AG, 2016), different types of data being collected in city scale digital twin settings like geographical data [P1 47:08/50:43; P2 33:13; P3 01:36/36:07; P7 05:07], traffic data [P4 31:41; P5 30:18] and infrastructure data [P8 01:57] introduces different considerations in data collection, storage and usage which are based on one's judgment or attitude as well as policy influence.

One interesting view from the responses is that regulation significantly affects the data handling approach. One expressed that GDPR is a good regulation but at the same time it creates insecurity and hindrance for them to collect and use personal data for technological development [P8 16:55]. Also, in order to avoid the risk of violating GDPR, personal data collection is mostly avoided [P2 51:54; P5 31:21; P9 09:13] as someone will have to be responsible if there is any violation of GDPR [P9 29:02] or anonymize the sensitive data before storage [P7 37:13]. The aforementioned view is explaining one of the accountabilities mentioned from a business perspective that they are obliged to law and regulations (Leese,

2014; Ananny & Crawford, 2018; Buhmann et al., 2019; Burrell, 2016) indicating the prevention of causing negative outcome to the organization.

A participant working in a public institution raised numerous ethical questions such as how much data they should collect, what is the limitation when collecting data, under what situation can they use the data and for what kind of purpose [P9 29:02]. The participant emphasized that those questions require clearer direction to guide them through the data handling processes in city scale digital twins, as it is more complex than GDPR since it depends on one's intentions [P9 31:22]. Those questions imply the complexity and dilemma of using data for achieving goals in city scale digital twins due to conflict of interest and agendas among stakeholders (Kitchin, 2016; Juvenile Ehwi et al., 2022; Helbing et al., 2021; Helbing & S'anchez-Vaquerizo, 2022; Dubber et al., 2020; Charitonidou, 2022; White et al., 2021). It is also due to the various accountabilities borne by different roles in the society, such as private, public or legal roles (Leese, 2014; Ananny & Crawford, 2018; Buhmann et al., 2019; Burrell, 2016; Ananny, 2015; Bovens, 2007). Particularly, a public institution is overseen by the government and working for the country, bearing the accountability that the actions they have taken are for the sake of the citizens (Bovens, 2007).

Moving to the technical point of view, it is important that people who are handling and utilizing data to tackle the societal questions in city scale digital twins possess integrity and reliability in reflecting the reality [P3 22:36; P4 27:01; P7 37:13]. It can be related back to the literature stating the responsibility for technical people to ensure the data used are accurate without algorithmic bias (Burrell, 2016; Ananny, 2015; Buhmann et al., 2019). As previously mentioned, there are insufficient guidelines or a lack of common standards that can be used by digital twin practitioners when handling data [P1 01:03:37; P2 07:53/13:01; P5 38:56/39:57; P6 01:15:01; P7 01:15:31; P8 07:02; P9 36:18/37:59]. Even though there is industry guidance on the data collection, it lacks explanation on where to collect data, how much to collect and what does not necessarily have to be collected, which is mostly based on individual's judgment and knowledge [P1 55:42; P2 07:53/13:01; P5 41:35; P8 07:02]. By way of illustration, two participants explained that the coverage of the data collection is higher in the city centre where most of the people live and activities are maintained, since there are more business opportunities [P1 11:48/13:07/58:54/59:01; P9 11:23/11:04] or research interest [P9 11:04], increasing the incentive to spend money on changes rather than prioritizing the outskirt areas [P1 57:34/58:54/59:01]. This perspective from the participants is echoed in the literature that data collection is more cost efficient when it comprehends areas with a lot of changes (Ferré-Bigorra et al., 2022), but this risk driving algorithmic bias as concerns of the cost and worthiness of an action results in selection bias (Akter et al., 2021; Stinson, 2022).

When it comes to the volume of data and in what level of detail data should be, participants had differing viewpoints which are unsettled. Data with higher levels of detail bring privacy concerns [P3 01:03:24] and security issues [P1 52:25], but it also supports better analysis [P2 23:18], also the extent to which an analysis is considered good or bad is dependent on people judgment [P2 26:29]. Furthermore, when the data is too detailed, especially illustrating future scenarios, people tend to falsely interpret it as facts, the participant pointed out a need to reduce the level of detail in data to prevent certain misunderstandings [P7 51:13]. It may come from the notion that people tend to believe data is stating scientific truth (Madsen et al., 2022; Brauneis & Goodman, 2018). The views mentioned hinted at people bias as the decisions surrounding data management is dependent on what they understand and interpret from their moral sense of viewpoint (Stinson, 2022), such as what is considered as good analysis for all and people's perception of data as the truth.

Although data can reflect facts in real life, it might lack the consideration in human values and societal aspects, which is hard to represent and measure in city scale digital twins (Helbing & S'anchez-Vaquerizo, 2022). Under the city scale digital twin, other than urban data, more abstract data is also being collected, such as well-being [P7 01:07:26]. It may be relevant to achieve goals enhancing wellbeing [P1 02:59/09:39; P3 07:28], but it also raises the questions of how it can be measured and represented [P7 01:07:26] accurately in digital twins and how well can city scale digital twins represent people behaviour or reflecting reality? For instance, traffic flow data aggregated to a yearly level [P7 36:49] or predicting pedestrian behaviour in different situations [P5 43:07] without using real-time or personal data due to GDPR [P1 51:32/52:25] that may cause extra effort to the organizations [P5 31:21/32:40]. When humans are presented as data, researchers urged practitioners to be aware of simplifying human's behaviour and values which may result in "an inhuman society" (Helbing & S'anchez-Vaquerizo, 2022, p.9). Especially when there are no common procedures or standards guiding practitioners in handling data and mitigation the risk of inserting human bias into it, the data accuracy as well as data representation may be compromised to some extent (Papyshev & Yarime, 2021; Ferré-Bigorra et al., 2022).

Participants expressed considerations of to what extent bias can be acceptable. As it depends on the purpose of the analysis, sometimes non-exact causalities are good enough [P9 40:48]. It also cannot contradict the value proposition and data trustworthiness communicated to stakeholders [P8 11:37]. Similar consideration is found from other participants as they think 100% accuracy is unlikely to achieve as well as cost inefficient; it can be acceptable if the outcome is being communicated, considering stakeholders will at least get a better picture of the entire scenario [P7 58:51]. The literature contrasts these views, as algorithmic bias risk bringing negative externalities like discrimination specifically against certain groups or people (Stinson, 2022).

Even though participants are not explicitly expressing the mitigation approach for algorithmic bias, we found certain measures are being used. First, other than the various accountabilities mentioned, participants suggested that practitioners as well as users who participate in city scale digital twins should possess knowledge on the data, especially how it is collected before usage or perform data quality check, otherwise they may not be aware of potential problem in data [P7 01:07:26 / 01:13:08; P6 01:09:54; P9 39:45]. For example, deciding whether a bus station is close enough in relation to a certain neighbourhood that is located on the other side of the train tracks [P7 45:02], it implies that if the data is collected before a tunnel is built, basing a decision on that data rather than more recent data will flaw the decision, differing from 4 kilometres to 80 meters in the example [P7 45:02]. This is an illustrating example connecting to historical data being used in the system for analysis, creating a risk of historical discrimination affecting future decision making (Stinson, 2022). They stated that educating practitioners is a strategy to enhance their knowledge for better management and utilization of data [P2 41:32; P7 45:02; P7 01:13:08/01:04:57; P9 13:35].

Second, common standards should be established to provide guidelines on how to maintain metadata and data versioning along with the responsibility towards data management regardless of its ownership in order to enhance data interoperability and data quality [P3 01:11:54; P7 01:13:08/01:04:57]. Zou and Schiebinger (2018) shared a similar idea that incorporating standardized metadata into data management processes can help users identify and mitigate bias as metadata is self-explanatory. While one participant held a counter view on this as they think metadata information may increase opacity as too many columns can be filled, it is difficult to

trace back the origin of dataset and as well as different variation on metadata format [P6 01:15:01/P6 01:16:00; P7 01:15:31].

Third, stakeholder engagement and citizen inclusion, to generate feedback or discussion to enhance fairer decisions, is suggested in the literature as another strategy to mitigate bias (Akter et al., 2021; Helbing et al., 2021; White et al., 2021). Although city scale digital twins contained various stakeholders in the city planning, there is only one project with citizen participation found from the participants' response in city scale digital twins project [P6 55:03/ P7 55:08]. Moreover, interviewees expressed the view that they intend to incorporate citizen feedback and participation in the future city scale digital twins [P4 07:39/01:02:12; P7 53:29/55:08; P9 27:44]. One of them pointed to the possibility of implementing digital twins to generate more interactive discussion with citizens [P9 27:44]. Another participant resonated to this notion as it is a democracy, everyone should be able to express their opinions and being considered [P7 55:08]. Besides citizens inclusion, stakeholder discussions are frequently found in the participants' responses, for example using digital twins to simulate different situations and scenarios to discuss the strengths and weakness as well as identifying bias for mitigation [P4 07:39/29:56; P6 46:57/47:31; P7 27:42].

5.1.1 Implication

Throughout the data management processes, almost all interviewees [P1 01:03:37; P2 07:53/13:01; P5 38:56/39:57; P6 01:15:01; P7 01:15:31; P8 07:02; P9 36:18/37:59] request common guidelines to follow as it will be the easiest way to achieve the purpose of managing data without causing any negative outcome. It is also the case that digital twin practitioners tend to avoid sensitive data or personal identifiable data [P2 51:54; P5 31:21; P9 09:13] to avoid negative outcomes such as violating the regulations. These are aligning with the consequentialism under the pragmatic domain to minimize the overall harm.

When they mentioned the people's value on handling the data, integrity is one of the virtuous values under the ethical domain. Besides, they are working towards societal questions in city scale digital twins to build a good life among the society [P3 22:36; P4 27:01; P7 37:13] which resemble Aristotelian and communitarian in the ethical domain as well.

However, from the example of more comprehensive data collection in the city centre compared to the outskirt area, it showed that they prioritize pragmatic domain with efficacy and utilitarianism to maximize majority benefits. Although the participant stressed that people involved in the project realize this and know that the goal is to comprehend all parts of the city [P9 10:37], this action introduced the risk of selection bias that may be detrimental to the minorities. It reflects to deciding the required volume of data and in what level of details it should be collected, certain values are mentioned such as privacy and security. If we simply look at privacy [P3 01:03:24], it can be considered under the moral domain as it is reflected as a fundamental right of the individual which is also universal. However, there can be a trade-off of moral value in order to obtain higher level details of data to provide better analysis [P2 23:18] that falls into the pragmatic domain of utilitarianism maximizing most people's benefit. But the analysis depends on human judgment on whether it is good or not [P2 26:29] which may possibly cause people bias that is unjust to the minorities.

Another decision on reducing the level of details to prevent misunderstanding as a trade-off of data quality or accuracy implies the consequences of preventing misunderstanding justifies the mean which is under the pragmatic domain [P7 51:13]

In the level of acceptance in algorithmic bias, we can imply from the responses that maximizing the greater benefits and efficacy, such as reaching a certain level of analysis [P9 40:48] and achieving cost efficiency [P7 58:51] are under utilitarianism while gaining customer acceptance [P8 11:37] is under the consequentialism which all are under pragmatic domain.

With different decisions and possible trade-offs within the pragmatic, ethical and moral domain, we did not see a proper TCA in place with all affected parties gathered for discussion. Some discussions were in place, but only internal stakeholders were included [P4 07:39/29:56; P6 46:57/47:31; P7 27:42/58:51]. However, we see the potential of citizen inclusion from the responses [P4 07:39/01:02:12; P7 53:29/55:08; P9 27:44].

5.2 Nudging

In the findings, participants reflected their views towards using data for nudging human behaviour in the digital twin city setting. For the sake of sustainability, it is not possible to explicitly tell people what to do but to inspire them by displaying more sustainable transportation options through the use of data [P4 55:58; P9 24:06]. Nudging behaviour becomes acceptable under the circumstances that the purpose is achieving something good for the society [P3 15:12; P4 55:58; P7 01:23:11]. However, there are different interpretations of good for the society as it depends on how people define and prioritize goals for the city such as efficient resource allocation (Yigitcanlar et al., 2018); enhancing city resilience; sustainability or human well-being (Pierce et al., 2017). In addition, literature stated the notion of "technological determinism", using technology to nudge humans changing their behaviour and lifestyle in a way to achieve the goal in city planning at the expense of human liberty and their freedom of choices (Dubber et al., 2020, p.5). While one of the participants provided a different view from the literature, they see data helping people to produce more choices than before in a more efficient way instead of limiting their choices [P7 01:23:11].

5.2.1 Implication

Nudging behaviour for the goodness of society such as sustainability is resonating with Aristotelian and communitarianism in the ethical domain. The positive outcome is to solve the grand issue that is considered as consequentialism. However, nudging may come with the risk of harming an individual's liberty which is one of the virtues in the ethical domain.

5.3 Privacy

In the literature, Kitchin (2016) mentioned that certain types of data like people movement data is considered as sensitive data as it tracks an individual's activity pattern and possibly generalizes them or makes inferences through algorithmic calculation, which causes the threat of predictive privacy harm, coming along with biases and discrimination. This privacy concern is explicitly expressed in almost half [P3 01:02:16; P4 38:37; P5 33:09 / 35:18; P9 34:18] of the interviewees' responses. When different datasets are combined, it is feasible and sensible for someone to associate certain relations through data patterns violating privacy [P2 51:40]. Other than individual's privacy, a participant expressed their worry towards data security at a

national level in relation to data patterns [P9 34:18]. When various collected datasets in the city scale digital twins are integrated, it may have a risk of exposing national security vulnerability as other countries may discover certain installations hold significant importance for national security through data trails [P9 34:18].

While one of the participants claimed that even though the movement data is measuring people, they are not identifiable [P2 51:40]. However, this notion is debatable as it depends on how detailed the data is being collected. Kitchin (2016) mentioned aggregated data is collected to show the "snapshot of cities at particular moment" which may echo the participant view that people movement data collected at a low data granularity level or being aggregated is not possible to trace back to one's information through their data pattern (p.2). However, when data-driven urbanism is implemented in city scale digital twins, the prevalent use of sensors and IoT technologies generate data with higher granularity level tracking people's movement in real-time which results in a richness of data being generated (Kitchin, 2016). For instance, using cameras with its sensing ability for data collection in the city setting can efficiently generate rich information for achieving a variety of purposes [P8 22:13] but it may also form a similar sensor society surveillancing the city with the risk of privacy issue (Helbing et al., 2021; Brauneis & Goodman, 2018; Kitchin, 2016; Petrova-Antonova & Ilieva, 2021; Juvenile Ehwi et al., 2022; Charitonidou, 2022). Thus, it comes back to the privacy concern as well as the predictive privacy harm issue, when more detailed data is collected, the risk increases of exposing people's identification through data pattern or being falsely generalized.

Even though some types of data may not entail explicit linkages to people it may still imply linkages with people through their data patterns or different circumstances. For instance, geographical data like zip code also hinted at the risk of violating privacy issues [P3 01:02:16] in connection to predictive privacy harm. Generally, zip code data does not show identifiable information, but it becomes problematic when the zip code belongs to a single person or business that follows the possibility of identifying them, even though it is not related to GDPR [P3 01:02:16]. People may reveal themselves and receive inferences based on activity related to that certain zip code. Therefore, some participants suggested that they would trade off the data accuracy level by aggregating the data in order to make it less sensitive or recognizable [P3 01:02:16; P5 33:09 / 35:18; P8 16:55] and ensure the privacy of people when achieving the purpose [P8 16:55].

5.3.1 Implication

As mentioned previously, privacy is one of the values from the moral domain. With increasing sensing capabilities to produce richness of data in the future, there is a potential trade-off of a moral domain like privacy in order to achieve city scale digital twin goal as in the pragmatic domain. While we imply the consideration in reducing data accuracy level [P3 01:02:16; P5 33:09 / 35:18; P8 16:55] to enhance privacy it is considered as duty stemming from deontology which falls under the moral domain, but it is debatable as lower data accuracy level may cause less accurate representation of certain groups of people. It will then become utilitarianism as pragmatic that only maximizes the goodness for the majority.

5.4 Data Sharing

Open data, releasing the data access to the public for common use is a recent topic being discussed in the literature (Dubber et al., 2020; Helbing et al., 2021). However, various considerations are found from the findings hindering the execution of open data. First, money is the prominent concern being mentioned [P1 18:01; P4 50:44; P7 35:08/35:47]. Since some data are procured from third parties or shared by business organizations [P2 41:32; P3 58:28; P5 46:22; P7 33:37/33:56; P9 15:44/32:27] instead of owning the data by the public institution themselves [P1 07:08; P6 31:31; P7 32:00; P9 15:44], they expressed the view that data has a value and it is debatable whether to share it publicly as open data [P1 18:01; P4 50:44; P7 35:08/35:47]. They can either sell their own data to sustain the source of income to procure or produce more data or remove the barrier of using data so that everyone can use it and create synergies in the city planning [P1 18:01; P4 50:44]. But the politicians do not fully understand the benefit of open data and ground their perception that tax money is being used for producing data and others extract the values from it, which in turn taxpayers bear the cost of open data [P7 35:08/35:47]. Their point of departure is mostly the money term while Dubber et al. (2020) stated that open data can prevent private companies from gaining value by exerting its power drawing from the data they collected in the society, resulting in reduced motivation to over collect data of the city. However, the second obstacle is that the data they procured or shared from the businesses have multiple stakeholders which they are required to obtain consent from in order to use or share, it can also stop them from sharing it since it is under a certain agreement [P1 07:08; P3 01:10:49; P9 17:32/34:18]. The third concern is rooted in the privacy and security concern as well as the fear of the data ending up being misused, violating laws and regulations, such as GDPR [P9 17:32; P3 28:35]. This concern is mirroring the business's responsibilities to protect their data assets, maintain competitive advantages among other competitors and abide by the law (Leese, 2014; Ananny & Crawford, 2018; Buhmann et al., 2019; Burrell, 2016). As mentioned previously open data is considered as public good which reduce the chance for the private companies manipulating the society or extract business benefits from the data used in city planning, it creates the resistance to them to as they cannot justify the possible benefits in sharing data openly in the digital twin city planning (Nochta et al., 2020).

Even though certain participants hold a positive and open view towards open data, it is difficult to balance the stakeholders' interest. We discovered a semi-open data approach in the finding. One use case stated that they provided ready to use data on the semi-open access city scale digital twin platform and let the related parties obtain data or information themselves which in turn streamlined the business processes enjoying efficiency gain [P6 35:01; P7 29:31]. The other one is that if the data is collected for one purpose, then the parties can use the data for free according to the specific regulation otherwise they will need to pay for retrieving the data for use [P9 15:44]. This semi-open data approach resonated with the twofold responsibilities that the role of the government is to uphold the public data being misused by the private institutions but also balancing the use of data from them to create a better society (Dubber et al., 2020).

5.4.1 Implication

Open data is considered under the moral domain as everyone equally enjoys the data access echoing to deontology [P1 18:01; P4 50:44]. However, the consideration is based on the pragmatic domain like efficacy on cost and resources [P7 35:08/35:47], possible negative consequences such as losing trade secret, data misuse or leakage [P9 17:32; P3 28:35].

5.5 Summary of Implications

To summarize the implications from our observation, we identified several activities where prioritization of the three domains were made within the data management processes and when encountering algorithmic bias in city scale digital twins. In total we found four activities where the pragmatic domain was prioritized, one activity when the ethical domain was prioritized, and no activities where the moral domain was prioritized. We also identified three possible trade-offs between the pragmatic domain and the moral domain, and one possible trade-off between the pragmatic domain and the ethical domain. There was only one explicit example of a TCA discourse process in the interviews. However, it included no citizens which means it was not comprehending all affected stakeholders in the situation.

The pragmatic domain is highly prioritized which implies emphasis on efficacy, within those activities utilitarianism and consequentialism are central, meaning "to maximize the greatest good for the greatest number and thereby risk injustice for the minority", and to "maximize the overall good or minimize the overall harm" (Mingers & Walsham, 2010, p.835).

Even though there is no independent prioritization of the moral domain identified from the activities, we found that the possible trade-offs between the moral domain and the pragmatic domain are more frequent than the possible trade-offs among ethical and pragmatic domains within the data management processes. As Stahl (2008a) stated "morality as the set of factual social rules or norms that are accepted within a certain group or society whereas ethics is the reflection and justification of such rules or norms" (p.4), we implied that certain ethical code of conduct should be developed in order to maintain the moral values such as privacy, in those activities. Besides, the findings also showed that participants called for common or even national standards in handling data [P1 01:03:37; P2 07:53/P3:01; P5 38:56/39:57; P6 01:15:01; P7 01:15:31; P8 07:02; P9 36:18/37:59]. We implied that practitioners require clear rules to follow when dealing with data, which resonates the idea in deontology of what is considered as right or wrong, fair or unfair.

6 Conclusion

The research question we are aiming to answer is stated as follows:

> How are pragmatic, ethical, and moral considerations prioritized by digital twin practitioners during data management processes, including collection, usage, storage, sharing and quality as well as when encountering algorithmic bias in those processes?

It was clear from the empirical endeavour that the pragmatic domain was prioritized higher during most of the identified activities in relation to data management processes and when encountering algorithmic bias in city scale digital twins. Four activities were found where digital twin practitioners prioritized the pragmatic domain, one activity where they prioritized the ethical domain, and no activities where they prioritized the moral domain. Possible trade-offs were identified between the pragmatic and moral domains, as well as between the pragmatic and ethical domains. While no independent prioritization of the moral domain was found, possible trade-offs with the pragmatic domain were found more frequently, suggesting the need for an ethical code of conduct to uphold moral values. Lastly, involvement of all affected stakeholders, particularly citizens, was lacking in the discourse process, hindering equal prioritization of all three domains.

6.1 Implications for Future Research

According to Mingers and Walsham (2010), when achieving democracy, it is required to fulfil pragmatic, ethical and moral domains respectively with deliberative discourse, using dialogue as engagement to reach consensus, when encountering the latest technologies and enhancing understanding of the rationale behind stakeholders' intention. Considering both the moral and ethical domain respectively needs to be further prioritized so as to balance all three domains under the democracy principle, in combination with the low density of TCA in the findings, there is room for improvement in this area. Further investigation of how to produce proper discourse involving all affected stakeholders in city scale digital twin projects could improve accountability and fairness in city scale digital twin projects.

Moving to algorithmic bias, digital twin practitioners are aware of this concept and discuss some aspects of it occasionally, some aspects frequently. Further research on how algorithmic bias appears in city scale digital twins, could aid those discussions and ultimately create awareness and standards to help mitigate algorithmic bias. Furthermore, this research is focused on the data management process and algorithmic bias in city scale digital twins. We propose further research on how city scale digital twins can incorporate data about human behaviour as well as human activities in order to accurately represent those social aspects in the digital twins and utilize it to comprehensively solve the social challenges, such as sustainability.

Finally, as multiple participants expressed the immaturity of the application of digital twins in the urban context, there is a need to define the specification of city scale digital twins in order for this technological approach to be further progressed. Also, several participants suggested to include politicians for discussion as they are significant stakeholders driving the city scale digital twin decisions, in order to discuss the possibilities and difficulties, especially how policy

is affecting the development of city scale digital twins. The comments hinted the need of cross disciplines research on the social, technical, and political aspect of city scale digital twins.

Appendix 1: Consent Form

Consent form

Master's in Information Systems, VT23

Thank you for taking part in the study investigating Algorithmic Bias in Digital Twin Design for Smart Cities. The study aims to understand the decisions made during the entire process of designing digital twins in smart cities as well as its outcome under the ethical perspectives. The implications of the study are to enhance awareness and provide insights to various stakeholders who are in relation to city-scaled digital twins development.

The study forms part of the compulsory program curriculum and is supervised by one of the assigned supervisors of the Department of Informatics at Lund University, who will ensure that students adhere to all the necessary rules. The resulting thesis is assessed and graded by an examiner at the end of the course.

We ask for your approval to use the data collected for the study. Participation is always voluntary. In order to collect data for the study, we need your signed consent on the second page of this form. Even in the case that you sign the form at this point, it is still possible for you to withdraw from participation at any time without motivating why.

During the course of the student working on the study your personal data are protected and will not be disclosed to unauthorized persons. We will store recordings and other details in a safeguarded manner. Any photographic/video/sound collected in the first phase will be anonymized, coded and transcribed as text. This will be done immediately upon transmission in order to disable any potential for detecting that you have participated. The consent forms will be kept safely so that they may not be linked to our recording. When the study is completed and the thesis has passed assessment, we will destroy the original data that has been collected.

The results of the study will be published in the thesis in a manner that will not reveal the participant's identity. The study adheres to the guidelines on research ethics and common laws.

In order to complete the study, it is very valuable for us to receive your consent. Please contact us in case you need further information.

<table>
<tr><td>Elin Wernholm</td><td>Wing Yi Lam</td></tr>
<tr><td>Email: el4610we-s@student.lu.se</td><td>Email: wi6457la-s@student.lu.se</td></tr>
</table>

Supervisor: Paul Pierce

Email: paul.pierce@ics.lu.se

LUND
UNIVERSITY

Consent Form

I have taken part of the information of the study and accept that the material is recorded and stored for use in the master's thesis.

☐   Yes

☐   No

The name of the informant, date and signature:

_____

Print name:

_____

Appendix 2: Initial Interview Guide

(City Planner = CP; Technical People = TP)

1. What is your role now, day to day work? (CP & TP)
2. Can you tell us more about the digital twin(s) you are currently working with in the city setting? (CP & TP)
3. What are the organizations, companies and people participating in the projects? (CP & TP)
4. What are the overarching goals of the project? (CP & TP)
5. How do you decide which areas to implement the digital twins? (CP & TP)
   a) Power relations based on funding?
6. What is the process of reaching consensus regarding goals to reach and problems to solve? (CP & TP)
   a) How are different goals weighted?
   b) Who is participating?
   c) How are people made aware of opportunities to participate?
   d) Are there any other groups or people you think should participate that are not?
   e) Are there any plans in action to include those potential participants?
7. What do you think is important when designing or implementing a system that is used for city planning? (CP & TP)
8. How is policy making affecting the development and implementation of the goals? (CP & TP)
9. Are there any limitations in the technology that challenge the reachability of certain goals? (CP & TP)
10. To what extent do you feel confident that you understand the technology embedding in the digital twin? (CP & TP)
11. Do you feel like you trust the technology (CP & TP) and the companies delivering the technical solutions? (CP)
12. Based on what kind of criteria do you choose the technical parties in the projects? (CP & TP)
    a) Contractors to technical companies? (TP)
13. Who owns success and who is responsible for negative outcomes? (CP & TP)
14. Looking back at the final solution, are there any discrepancy you find from the goal you set at first vs now? (CP & TP)
15. What kind of data is collected for the different goals of the digital twin? (CP & TP)
16. How do you collect the data in order to achieve the goals that you mentioned? (CP & TP)
17. How are the groups and areas affected by the digital twin actually being represented in the data? (CP & TP)
18. What is your current understanding of algorithmic bias? (CP & TP)
19. Is bias something you discuss within the project in relation to the digital twin? (CP & TP)
20. As an example: uneven distribution of the data points collected might result in bias, for example more data points being collected in the city center than in the suburbs. (CP & TP)
    a) How would you want to manage that within the project?
    b) Are there any standards to deal with it?
       • General or company-specific?
21. To what extent do you think that the digital twins model is simplifying its real world activities and complexities of human activities? (CP & TP)
    a) How are you managing this fact when making decisions around city planning? (CP)
22. What is the industry standard for collecting data without citizens consent? For example through sensors or cameras etc. (CP & TP)

23. Under which circumstances do you think it is appropriate to collect data without consent? (eg. through sensors) (CP & TP)
24. Who owns the data? (CP & TP)
25. What is the process of governing data usage? (CP & TP)
    a) Who/How?
    b) Disclosure to the public?
26. How might the digital twin have an impact on the citizens ability to make their own choices for their behavior and lifestyle? (CP & TP)
27. What is the consideration when data collected for a certain system is transferred into other systems? (CP & TP)
    a) Purpose?
    b) Consent?
    c) Bias? Reinforcing over time
    d) Goals?
28. Do you know if there are any processes or standards to prevent biases, personal opinions or preconceptions from being built into the systems and how it affects the data quality? (CP & TP)
    a) Are those standards applying to all technical people? (TP)
29. To what extent can bias be acceptable in processing the data? (CP & TP)
30. Are there any further limitations/concerns that you can see that we did not discuss yet? (CP & TP)
31. Is there anything else you would like to add that we did not discuss yet? (CP & TP)
32. Do you have any questions for us? (CP & TP)

Appendix 3: Interview 1, P1

Theme	Colour Codes	Factor	Text Codes
Context	Yellow	/	/
Data Management	Blue	Data Collection	DC
		Data Storage & Use	DSU
		Data Sharing	DS
		Data Quality	DQ
Algorithmic Bias	Red	Data	BD
		People	BP
Principle of Discourse - Democracy Principle	/	Pragmatic Domain	PD
		Ethical Domain	ED
		Moral Domain	MD
		The Theory of Communicative Action	TCA

Participants	
Authors	SPK_1
P1	SPK_2

Transcript	2nd layer	3rd layer
[00:00] SPK_1		
On both our computers just to be safe, in case anything happens.		
[00:04] SPK_2		
Are you are used to listen to yourself speaking?		
[00:08] SPK_1		
No,		
[00:09] SPK_2		
it's a <explicit expression>		
[00:12] SPK_1		
Oh. That's gonna be an experience		
[00:18] SPK_1		
So, okay. Yeah. So would you want to start telling us a bit about your role now your day-to-day work?		
[00:24] SPK_2		
Sure. My name is <Name>. I'm <Role> at <Company Name> in <Location Information> Sweden. My role is particularly I deal with a lot of data data management in various situations where data is comes from some place and should	DQ DSU	

go to another place but the structure isn't that good. So, you need to restructure the data and so that's where I am. I usually be between if you draw like a like a sketch like a box of the arrow and then another box. I'm the arrow between the boxes that could be like two different systems, but it could also be two different disciplines within an organization and the for the last two years. I've been a deeply involved in a project with <Location Information> where they try to invest how investigate what information model for digital twin or like a 3D City should be like and I was the one and doing that job together with them. Yeah,	DS	
[01:40] SPK_1		
so that's the project that you're currently working.		
[01:45] SPK_2		
And ah, but		
[01:47] SPK_1		
is it multiple projects or is it just		
[01:50] SPK_2		
yes, a <Role> would do what we where we are needed to		
[01:55] SPK_1		
OK.		
[01:55] SPK_2		
And sometimes I can do just one project the one sometimes I do several other different projects in parallel. So, a day-to-day basis my my calendar look a lot of different. I can be in a meeting with a client then on the having lunch and on the the second half of day I can sit with a data problem and try to solve it. So, it it's very it's variety of different tasks that I need to perform each day or each week.		
[02:25] SPK_1		
Okay,		
[02:26] SPK_2		
but but you're studying in Lund or		
[02:30] SPK_1		
yeah, we're studying in Lund. Yeah, Lund University.		
[02:33] SPK_2		
Is it some sort of like a program or		
[02:37] SPK_1		
yeah, Information Systems a master program one year master program and information systems.		
[02:44] SPK_2		
Okay. Yeah.		
[02:46] SPK_1		
yeah, and so the digital twins that you or twin that you've been working on in the city setting. Could you tell us a bit more about that?		
[02:59] SPK_2		
Yeah. Well, when it's think about digital twin? It's a hard. It's a hard thing to Define. What is actually a digital twin of something? Is it one to one twin with all the details from the the physical world to the digital world or what is it? And we thought we should go into that hole that rabbit hole because we won't get out to Define what things are. So, what we looked at was. hmm. From the geospatial side because I'm a <Role> in the geospatial the JAS geographical point of view of all these data that the city collects what can we reduce and how should that be defined so we can reuse it in a digital twin no matter what a digital twin actually are it could be a digital twin that you could Like it test different scenarios before it they build the city. Like noise	DC DSU DS DQ	

if you build a certain like a new Street and you build a new houses are they affected for noise from the traffic. can we can we have the the buildings in another way and so that the noise won't affect so many residentials. So, and how can the data from the city be the foundation for such an analysis? What need the data to be like? How can we collect the data? How can we store it? How can we like and that exchange the data we got so we can do this cool scenarios that in the future. So that was like the aim because when we come to a city, they have collected geospatial data. For hundreds of years. First, they did it in paper and then some way in the digital world and then they try to store the the paper map in the database. So, it's just lines and dots. But now we are transitioning to be how can we store 3D data like volumes and surfaces and how can we store that and connect these to each other? So, a wall so house know where on the surface that situated so the surface aligned the walls a lot of technical questions. Yeah, so that's where I've been the last few years. Investigating needs and try to take those needs write some sort of definition of how data should be like Something like that.		
[05:53] SPK_1		
for the <Project Information> projects that you mentioning. Is this for all the city that you mentioned like <Location Information> or just for one city.		
[06:04] SPK_2		
It's it's not just the project. It's like the data that need to be structured in a way. So	DSU DQ	
All the cities can do such analysis.		
[06:13] SPK_1		
Okay. I got it.		
[06:14] SPK_2		
It could be noise. It could also be where is the best place to to place solar panels on the roofs? And how will share those affect different apartment in? In a city such analysis which we as. Like like we will think that's all it always happens that these such analysis but they don't because they are very heavy of collecting data and need to store the data and it's a versioning that data and how on Earth do you do that for like a city where we have thousands of buildings?	DQ DC DSU	
[06:56] SPK_1		
Yeah, so that's what you're working on.		
[06:58] SPK_2		
Yeah.		
[06:59] SPK_1		
Okay. Nice. Yeah, so we were wondering what kind of organizations and companies and like people are participating in these projects.		
[07:08] SPK_2		
Well, they all often drive, hmm, establish from the municipality because they are the one who having the data. and the data that the municipality, the official municipality, they they are thinking of having out this data as open data free to use and then the society like the build contractors and byggherre in translator in Swedish to English the ones who construct the architects that will construct new houses. They will take these data from the municipality and use it. So the things that the municipality then get back from the architect is in a better way so they can take wiser decisions. So my projects are often driven from the official side the municipality or the governmental side of things and I think that's the way it should be in in Sweden because we are like such a big public administration that all the cities they can do whatever they want. it's not like a private company that owns	DSU DS DC	

data because that will It will keep a barrier from using the data if you need to take some sort of payment for it.And the data that should be the ground for these analysis should be collected and given out from a public side. So Long long answer to short question.		
[08:46] SPK_1		
Yeah, so what would you say like in the city setting with this digital twins are the overarching goals of the project like wide		
[09:04] SPK_2		
well, I think the major. the major thing is that the municipality has done the same thing for a long time and now they are changing the way to doing things and that's really hard for the organization and for the people working there to change a mindset of how they do things and why they do things and I don't remember the question actually		
[09:35] SPK_1		
what what are the overarching goals of the		
[09:39] SPK_2		
The overarching goal the overarching goal is to be more efficient to try out things. Before constructing it. But if you if you constructing a new like a new block of a square or block of a block of houses, and then when you build it, you understand that we cannot use this houses because a lot of noise pollution or they are very dark then it's just stupid. But if you could try it out in a digital environment first, and then you can find the best settlement for how the house that should be and then it's the overall the goal would be to make wiser decisions before constructing them. That's one goal.		
[10:26] SPK_1		
Yeah based on the digital twin and what it can show you.		
[10:31] SPK_2		
Yeah exactly.		
[10:32] SPK_1		
Yeah,		
[10:32] SPK_2		
and it can also be now we're talking a bit of buildings and houses, but it can also be when you have a lot of perception from rain and where the the blue spots are where all the the water will collect so we can construct the city with that in mind before the rain occurs the hundred year event where all the flooding will do and then we have like an analysis of where we should do certain things will affect how we construct to say this so we can avoid these flooding situations. Yeah.		
[11:11] SPK_1		
Yeah for crisis situations		
[11:14] SPK_2		
could be that it could be anything exactly.		
[11:16] SPK_1		
Yeah.		
[11:19] SPK_1		
And how do you decide maybe that's not you but how do you know how it's decided where which area is to implement The digital twins?		
[11:33] SPK_2		
What is speaking of areas? Can you give like an example of what areas you're having mind?		
[11:38] SPK_1		

And I'm thinking if there's like there are specific neighborhoods or		
[11:43] SPK_2		
okay, like the geographical areas. Oh, yeah.		
[11:48] SPK_2		
I don't. What hmm..where okay, I can give it an answer to that actually because You want to aim your your power when constructing the digital twin to somewhere? Where they're gonna be a lot of changes or well the the change will affect a lot of people. So if you take like a municipality out in nowhere in Sweden, and they have a lot of forest. It's not that wise to start there with a digital twin because nothing will happen for like 200 years. But if you're in the city center and you can make you have to demolish a house and build something new that will affect a lot of people a lot of people will have a lot of ideas of how and why and the the city could should do then digital twin could help you with these decisions and with this information campaign to the residents. This is how we will how did the new house would look like and why it looked like this and what other underlaying problems we are dealing with so city center. A lot of changes that will affect a lot of people.	DC BP BD	PD
[13:12] SPK_1		
Yeah, so you're using the digital twin and to communicate Changes and		
[13:17] SPK_2		
there's also a goal with this to communicate changes and get an involvement with the society. So they understand what's going on why it's going on and also to to retrieve a lot of ideas from the from the society what they think this city should do and why for the could be like we construction this new block of houses, but you then society say that this is uncertain I'm feeling unsecure here. It is very dark. Can we in some way and make it more light or could it be some the environment get more secure and that would they say they will do that as well. It's part of the project. So communication with the societies are also important. but speaking of that right now, there's not It depends on what you say at the digital twin is if a digital twin is a 3D model of the city well, and it's okay. This is just a 3D model we deal with City then a lot of cities have done it. There's a lot of citizens in Sweden who have a like a 3D model online when you can look at how the city's look like and what it should look like but the 3D model that it Which is connected to like IOT sensors where you can control traffic flow and you can make a lot of different scenarios regarding things. I don't think any City in Sweden has that yet?		TCA
[14:52] SPK_1		
Okay. Okay. So it means like that because what you say that so it means that there's no Central brain to control the data. So instead of just sending the data and then analyze it and then you make some prediction or like simulations things.		
[15:13] SPK_2		
They could be like blockwise or like project wise analysis right now. We're building a new house over in <Location Information>. They want to do like a 3D model they want to to see you if it's good situation for solar panels and things and then you do it just for that small area that they have another project in the same municipality and do the same analysis could be on the same data source. But the data analysis isn't stored anywhere.	DSU DS	
[15:46] SPK_1		
Okay,		
[15:47] SPK_2		
and you cannot see these things in parallel. They are like small stamps Of information	DS	
[15:54] SPK_1		

understand. Yeah.		
[15:58] SPK_1		
Okay. Yes, so the next question what is the process of reaching consensus regarding like goals if there's a contradicting goals or		
[16:14] SPK_2		
with creating a digital twin?		
[16:12] SPK_1		
Yeah and projects in general like if		
[16:15] SPK_2		
oh, yeah money.		PD
[16:18] SPK_1		
Yeah, okay, that's interesting. So if there is a lot of goals then how can you like weight which goes to go first? So later how you do the..		
[16:30] SPK_2		
It's hard for me as a <Role> to answer that question. Actually, that would be more appropriate if a municipality or city can answer how they deal with multiple. Disciplines that have different demands but what I see is that hmm. if something is very expensive and doesn't make the big change then you won't do it so they try to find. The most bang for the money like how could we do things as much efficient as possible with a least amount of money spent?		PD
[17:12] SPK_1		
mmm		
[17:13] SPK_2		
Which is kind of hard because you don't know before		
[17:16] SPK_1		
understandable. Yeah.		
[17:20] SPK_2		
It's hard. Yeah, actually.		
[17:23] SPK_1		
I think then there could be some kind of power relations based on funding and who has the money in those decisions or		
[17:33] SPK_2		
Yeah, well. an interesting thing here is this is a political decision because the money where the the city got their money from two different sources or okay, the major source is like taxes from the the society on the inhabitants and the politicians don't want to raise the taxes.		
[18:00] SPK_1		
No.		
[18:01] SPK_2		
So, how can we keep taxes low? Okay. Another thing is like selling data. a lot of municipalities sell their data for money If you make something open like an open data open source, then you cannot take money for your data. Then you lose that source of income. So will you raise the taxes? To spend the money of giving data as open data. What happened to the the people working there? Can they are they allowed to work there? Does the municipality have the money to keep them employed? There's such a different aspects of the of the question but it's actually quite interesting how the politicians and their politics control the way the municipalities do things. We're only comes to digital twins when it comes to the open data the question and stuff. So a lot of these things we're talking about now is things that cost money for the municipality or the city. So we need to find a way to	DS	PD

save money. Mm-hmm, but save money isn't the same as Have you cannot quantify save the money in that way? It's really. Because time isn't money in the municipality. The time that the employees spend of doing things when you are employed at the city isn't quantified. It doesn't cost anything. So this is kind of hard. So it will be really interesting for your thesis to have an interview with a politician or something like that and talk about okay. Where should this? A lot this cool 3D digital twins things. Where will the money come from?		
[20:10] SPK_1		
Ok so the money is a big		
[20:12] SPK_2		
money is a big question. Yeah.		
[20:14] SPK_1		
Yeah, okay. Yeah, so because we were thinking about who is participating in these.		
[20:21] SPK_2		
Yeah		
[20:22] SPK_1		
decisions but that's then based on. Money or would you say that? Like who's who is included in the discussion		
[20:30] SPK_2		
the discussion about the digital twin? Or the discuss. Yeah		
[20:36] SPK_1		
Yeah or reaching consensus about the goals or all conflicts you want to solve or something like that?		
[20:48] SPK_2		
I will say that most of these projects has been established from the data point of view or the technical point of view. Not the Not the discipline at the municipality who got the most. earnings or because they perhaps when they when they when you live in a city and you're doing like yeah, let's take an example. Cleaning the streets or you're managing the people who clean the streets then perhaps you don't understand the possibilities with the technology. That the technology could look at the data look at your streets and predict when you should do when you should clean the little boxes. So you don't have to go there when they are empty which is kind of stupid. But the people working with these questions doesn't know technology and the possibilities. So a lot of these projects are like established from the technical side and the technical side doesn't have actually the use for the the solution. They just interesting of hey, this is a solution for you. You should be more effective so they will be more like a project internally within the city where the technology side need to like sell the solution or the idea to the other side within the same organization. So it's it's hard. It's really hard to to see where the goals are and how you should like do a rating if this goal is more important than that goal because people aren't that into the question yet.	DSU	PD
[22:44] SPK_1		
Okay,		
[22:44] SPK_2		
sorry.		
[22:47] SPK_1		
No, it's an interesting perspective that we got.		
[22:50] SPK_2		
This is the first interview to get.		
[22:54] SPK_1		

Yeah, yeah,		
[22:56] SPK_2		
it's gonna be really interesting to see what other people say.		
[23:00] SPK_1		
Yeah now so I read some articles so it's different perspectives from the scholar and from your view. So I think it's very interesting to hear about that.		
[23:08] SPK_2		
Yeah.		
[23:09] SPK_1		
Yeah. We want to ask also about the citizens here. I think like are they included somehow in the discussion or?		
[23:22] SPK_2		
Mmm, not in my project not yet. They <u>will be included</u> in the way that they got used to digital twin to to have idea at the sense in synpunkter (point of view) now.		TCA
[23:36] SPK_1		
Now their point of views		
[23:38] SPK_2		
Yeah their point of views of certain project where where which is visualized within the digital twin but no, <u>they are not included</u> in in my project not		TCA
[23:47] SPK_1		
but for the solution, will they be like the one who testing the solution?		
[23:53] SPK_2		
In some way. Yeah, that would be the end user. <u>Everything I see it does is for their citizens So the end user will be the citizen, but but perhaps the citizen in the everyday life doesn't use the solution because they can't see the use of it because you always go to Google maps to find things.</u>		TCA
[24:18] SPK_1		
Okay?		
[24:19] SPK_2		
Yeah. But it could be like if you have <u>a digital twin within a city you can do predictions. The city itself could do predictions when they do like. Yeah cleaning the streets where a lot of to avoid the noise pollution sort of thing.</u> Then <u>the citizen will be affected that but in an indirect way.</u> Because you don't feel oh, I'm a citizen. The city are done a really good job emptying the litter boxes here. You are happy if they are empty.		
[24:57] SPK_1		
Yeah,		
[24:57] SPK_2		
and then it doesn't you don't even reflect that. They are empty.		
[25:01] SPK_1		
Okay. Okay, so would you say that the goal is kind of for the citizen not really to notice that like just to		
[25:11] SPK_2		
make the life better for the citizen.		
[25:13] SPK_1		
Yeah, and		
[25:14] SPK_2		
life better will be avoid traffic jams avoid noise pollution get the information as easy as possible when I need it. And pay as little as possible in taxes.		
[25:31] SPK_1		

Yeah, that's a hard equation.		
[25:34] SPK_2		
Yeah, it really is but if you take it yourself. You don't need if you if you if you cannot sleep at night worrying about to meet the city and how they dealing with data. Then they've done something wrong. Did it just work for you?		
[25:50] SPK_1		
Yeah,		
[25:51] SPK_2		
and as a citizen this should be as easy possible to do things. As a citizen in your city. And how is this done? With the help of digital twin, I don't know. We're not there yet. No,		
[26:07] SPK_1		
okay. Okay. so this is a pretty broad question and What would you say is the most important thing when designing and implementing A system that is used for City Planning.		
[26:22] SPK_2		
Yeah.		
[26:26] SPK_1		
It's your opinion.		
[26:28] SPK_2		
Yeah, that's a good question. I think the most important thing is. you need to build something that you can reuse the data and who should be responsible for? Making the data up to date who are the one in charge of the data. Because a lot of things we see from <our> side of you is that city spend money and time doing things. That when they are finished they look absolutely superb. It's really good 3D art and beautiful. But as soon as the city or someone started demolish a house or reconstructed house, then the 3D model is out of date. So, how can you make things live throughout the years? Which means how can you collect your data? Analyze your data sort of Hantera	DSU DS DQ	PD
[27:34] SPK_1		
manage		
[27:36] SPK_2		
manage your data. So it's alive and who are the one who say I'm responsible for the this data. So that's the thing. You're still figuring out in the projects. You're well. It's kind of hard because nobody wants to be responsible for fix.	DQ	ED MD
[27:57] SPK_1		
No. Okay.		
[28:04] SPK_2		
but it's really hard because when we speak about the 2D printed Maps, which we all have we know of like a paper map the same day as you print your paper then you know, it's out of date because something has happened. Yeah. So, how do you think keep data up to date? It's really really hard.Who should do do we need to have people out in the city to see what's happening and then update your data. And if you have like a hundred different robots out there doing things just overlooking at what's happening in the city who should pay for them? So right now the cities are kind of responsive they get they got oh something has happened. We need to update the data.	DQ	MD ED PD
[29:00] SPK_1		
Yeah,		
[29:01] SPK_2		

but something can happen like three years ago, and then you have data which is three years old.		
[29:12] SPK_1		
are you working with sensors or that kind of technology or		
[29:17] SPK_2		
it sensors are they are incoming in the cities, there are not so They're more of like a sense of traffic sensors and environmental hazards like pollution and such things. And these sensors are really good to to make decisions. Both predict things what will happen in the future different like weeks or over in the year, but also take decisions there and now like, okay now the levels of pollution is very high. We need to close that street. Whatever. but when we speak of a sensors a sensor cannot like replace everything that the municipality need to do sensor itself cannot populate a 3D model of a house, which is a refurbished.	DC DSU	PD
[30:19] SPK_1		
Okay. So other than the data is not like up to date enough. So is there any limitation in a technology that is kind of like hindering or give you challenges of reaching the certain goals when you're developing the solution?		
[30:35] SPK_2		
Hmm yeah a lot of different. One thing is if you take a house and then you what should the cities store when it comes to house? Which level of detail LOD is a really good. Way to spend all your life figuring out the definition of but and if you come to a house is the 2D is the 2D footprint of the house the city should store. Is it a 3D box of the house? Is it a 3D representation with the walls the slab the different floors and the roof? Are there different geometrical objects? Or is it do we need to store the interior as well? What should we do with a windows? So defining the level of level of detail that the city should store how they collect that data. How do you do versioning? Like control the different version of the the house or the street. These are really tricky Parts technical side. Because you need to store things in the database, but the usage of the data could be anything from a 3D software to a 2d map to like. A cell phone or whatever the need to like consume these data in different ways.	DQ DSU	
[32:16] SPK_1		
So is it more about the decision of which type of data and how? Like decisions around it or is it actually limitations in the technology? That's		
[32:27] SPK_2		
I will say that the most tricky is part is the decisions around the data.		TCA
[32:33] SPK_1		
Yeah. Okay,		
[32:35] SPK_2		
you should look into <Location Information> in their digital twin actually and they have a lot of ideas and conceptual models of how things would look like.		
[32:49] SPK_1		
Ok. Yeah <Location Information>.		
[32:53] SPK_2		
Oh great.		
[32:53] SPK_1		
With some answers. Yes.		
[32:55] SPK_2		
They are really really good to speak to.		
[32:58] SPK_1		

Yeah, okay. That's nice. And and you work from the technical side. but you do you feel confident that you understand the technology that is included in the digital twin like		
[33:17] SPK_2		
both yes, and no because the definition of a digital twin If you start with a word where we're talking about, that's hard. That's I speak to a lot of municipalities and say oh we were playing of establishment digital twin and I was like, so what is a digital twin and nobody can answer that and that's a really hard thing because when you say digital Twin and <Name> said it to Twin and I said did that when we have a different perception of what that is and then if we're talking to each other are becoming to the to the like discussion with different point of views what this actually is then. It's kind of hard to to discuss things because we understand things from different ways. So if I come from a technical side of you I know a lot of data that sounds really stupid to say, but I know that the formats and how data are collected how they are stored and So that's not an issue. It's more like the different words we're using and I will say	DQ	
[34:29] SPK_1		
the consensus of the terminology and		
[34:32] SPK_2		
yeah,		
[34:33] SPK_1		
yeah.		
[34:35] SPK_1		
And is that somehow being worked towards? Like are you trying to solve that discrepancy between or		
[34:45] SPK_2		
I think why a widespread solution to this is just skipped that part and go directly to the question, which I think is a bit sad. I think it should be like a definition of what what is actually we're doing dealing with another thing that we are dealing with is that let's say you got a particular example you have like a building legislation in Sweden where you need to have a permit building permit to be the house and a lot of architects are using 3D software to decide the construct your house in the 3D environment where you can set like how the sunset will look in the third floor on <Date>. The technology is there but when they had in this to get the building permit, they need to print this as a PDF. And the decision is made on the PDF and goes to archive. Which means that you do if you do a lot of intelligent work on a computer and then you kill all the intelligence to PDF.		TCA
[36:10] SPK_1		
Yeah		
[36:10] SPK_2		
that you build the house. And then the same municipality go out in the field. And collect the data where the house is. So you get something smart. There's something stupid and then you're trying to to make that smart again. So I think a lot of and why this is why is this done this way? Well, the law in Sweden says that you cannot take a decision on this digital 3D model. You need to have it as a paper. I think a lot of legislation on the laws need to be reconstructed so we can have a whole digital 3D workflow from ID to management from the permits the construction work and then the management of the constructive building.	DC DSU DS	
[37:09] SPK_1		
okay, so the policies around this are taking away the intelligence		

[37:14] SPK_2		
definitely definitely they are. slow to and to adapt which is kind of good in one way, but it's a bit distracting from what we actually can achieve with the technology.		
[37:32] SPK_1		
Okay. And yeah. so we have so many questions left based like when you choose contractors and things in the project, are there any criteria that you're following		
[37:55] SPK_2		
Yeah so that question will be if I was a municipality and I		
[38:03] SPK_1		
Or will you have other like technical people that is your contractor to help you guys to build a solution together or you actually is in-house solution just built by <Company> solely.		
[38:13] SPK_2		
we building them. Well what we say that we are like vendor independent that we can choose which vendor software we we think it's appropriate for the situation and what our if I work for <Location Information> municipality. They decide what I should use.		
[38:36] SPK_1		
Okay,		
[38:38] SPK_2		
so they say, okay now you should only do things in the <Company> software.		
[38:44] SPK_1		
Okay		
[38:44] SPK_2		
software part that I need. Okay. I need to find make the solution just <Company> based.		
[38:52] SPK_1		
Okay		
[38:54] SPK_2		
that could be one thing? Yeah		
[38:55] SPK_1		
is that because they are basing their selection on like open like they're putting it out in the open and people get to give their Under there's or like there's some kind of process around that right? Which		
[39:13] SPK_2		
oh, yeah. Well, that's first of all, yeah a public. Beställare come to a public decision maker like all the municipalities since we don't they cannot choose.		
[39:31] SPK_1		
No,		
[39:32] SPK_2		
there's no shoes. they need to within the law find do a procurement to find the best. And solution partner contractor consultant to solve their task.		
[39:52] SPK_1		
Okay,		
[39:53] SPK_2		
they cannot choose which means that the procurement like where they they can end up with something. They don't want. but they need to use it because that is what they procured and got. And that could be on the software side like, oh, we we have this platform of GIS tools or data management tools. We don't like them we		

hate them but they were the cheapest so we got them and now we are stuck with them for five years.		
[40:39] SPK_1		
Yeah. Okay, so based on money as well.		
[40:44] SPK_2		
the whole law around procurement in the official side is you need to take a decision on quality and or money, so it depends on. sorry <Name>. I will switch to Swedish now for a short while. I will switch to Swedish now for a short while. Upphandlingsenheterna på kommunerna är dem som bestämmer hur upphandlingen ser ut och beroende på hur dem har formulerat upphandlingen så får dem svar, och dem ska utvärderas på pengar och kvalitet och relationen mellan dem. Which means that		
[41:22] SPK_1		
the money and the quality		
[41:24] SPK_2		
the money the quality Universal words		
[41:26] SPK_1		
Okay,		
[41:27] SPK_2		
which means if I'm municipality say, okay, we need a software for a 3D digital twin and it should be the lowest price will be the winner. And then you have a certain criteria that okay, the software should do this. They should do this should do this. And if you come with your software, which can do these things. And it's really really cheap then they are the winner, but then it turns out that it's not that user friendly and the municipality or city has forgot to take that. There should be a manual or they forgot that the criteria that they should have a graphical interface and then you can perhaps stuck with something in Bulgarian and you can all only call it from the <Company> prompt like with commandos. So be it. Which is kind of hard and that's why we see from all 290 cities or municipalities in Sweden. We see different technical softwares together because they have done 290 different procurement and they've got 290 different answers. So it's 290 different environments, which means that one solution that works for one municipality doesn't work for the other.		
[42:48] SPK_1		
Yeah so do you think that's a limit like a big limitation of the in the projects to to manage those kind of different solutions and		
[43:01] SPK_2		
I think it's hard for a small municipality to to take benefits of what other municipalities has done because you can't take their job and just implement it in your own municipality because the software environment will look a lot different so when coming from the vendor side as we are or because the the private side and <our> side we see that it will be a benefit as everybody do the same thing in the same software, but if you come from a software side and say well as a suit hopefully this is my software that make everybody can do the same. But if this is another one, then there's not Okay, so I think we just need to deal with this. We need to find the the core what we can reuse in the different municipalities and if that is like a data model or an information model of what data should look like what which data and how we should look like and then all the software vendors need to adapt to that perhaps that's a solution.	DSU DS	
[44:11] SPK_1		

Okay. So for the solution that we are talking about a lot. So who owns these success of the entire project or do you think like who is responsible for the negative outcomes that from all those projects solutions?		
[44:28] SPK_2		
Well, it's all these 290 different municipalities or cities. They are the ones in charge. We are actually seeing now the national mapping authority. The national land surveying launched material yet in Sweden. They are they're trying to do like national standards of how data that could be. The foundation for digital twin should look like the only downside to this is it take a lot of man hours or women hours for that matter as well and to come to the conclusion of how things should be like with speaking of like years of years of years and the things that they decide or come up with like this is the way data should look like It's not mandatory to follow you could do it but you can also do it as is as you want. and that's like the foundation from the swedish way of dealing with with decisions that the municipality they have their their own rights to take their own decisions of how things are done.	DQ DC DSU DS	ED MD TCA
[45:45] SPK_1		
Okay.		
[45:46] SPK_2		
Kommunalt självstyre, but it's lovely.		ED
[45:50] SPK_1		
Yeah, okay.		
[45:53] SPK_2		
So if you were like a dictature done that the state will say that everybody should do like this. You need to follow otherwise to go to prison, but that's not the case here. This is more like well, it's really swedish you we think you should do like this. if you want		ED MD
[46:15] SPK_1		
So the decision making is a hard.		
[46:18] SPK_2		
It's really hard. Yeah.		
[46:21] SPK_1		
Okay. I'm thinking if we should go a bit more into the data now. it's very fitting I guess since you are work with that a lot		
[46:32] SPK_2		
Yep.		
[46:32] SPK_1		
We don't have much time left. I don't know. Do you have another meeting after or do you have		
[46:37] SPK_2		
no I can spend some more minutes on this to at least until yeah quarter to two something that yep.		
[46:47] SPK_1		
Thank you. Yeah. Okay. So we we go into the the data then I think we already answered some of these questions. Or Yeah, yeah, because because we are the first question we want to ask what kind of data that you collected for the digital twins, but from what you are answering so it's more about the spatial data.		
[47:08] SPK_2		
Yeah		
[47:09] SPK_1		

That collected but other than the spatial data, is there anything that you collected to supplementing the digital twins?		
[47:16] SPK_2		
Well, that's interesting question because just if you just take your spatial data, like the things that we can see with our eyes that the the digital twin will be kind of boring maybe like road, fences, yeah house. Yeah. But so what we are thought of in our project with the <Location Information> is okay. How can we collect the the things that we see? But all the other things about the things we see like if you take a tree then we know we can say where the three situated. How big is it? Sorry. I need to cough	DC BP BD	
[48:00] SPK_1		
Okay.		
[48:03] SPK_2		
I can see it where the the tree situated I can see how the diameter and where the the leaves are how high the tree is. But there's a lot of different information about that tree in the municipality. It could be with sort is it. in both like the Swedish and in Latin? When was it planted? When they have you cut it, how did it had some diseases? How shall you cut it? When should you water it? Who owns it? A lot of different questions, but that that data is more like the data that the national the park organization in city that will deal with the data. So what we think in the digital twin is that we should only connect that data. We cannot do a data model with every information. from everything in the city that data model will be enormous and you cannot implement it anywhere and it will be a pain in the ass to it like export some data because it's really like huge but we try to like find how can we collect and store where the tree is and how can we have some sort of ID so we can connect that three to the park database where all the information about this tree is that's the connection. So and we have thought of this for buildings for roads for like what we say as city furniture like things that are within a city with which are man-made. And also like plants water bodies. Yeah, listen, that's a standard called City GML which is worldwide used for like digital twins inside the city, which is we have looked into and see can we reuse it? And what shall we add to that which is missing from the standard?	DC DSU DQ BP BD	ED
[50:22] SPK_1		
Okay. Sorry. I got a really stupid question. So for the data, so you means that actually the data is already there and then you guys like more reuse it and then connect the data how it fits to the goals that you want to achieve. This is something like that or you will supplement with some data that you collected by your company yourself.		
[50:43] SPK_2		
My company doesn't collected data.		
[50:45] SPK_1		
Okay, so		
[50:46] SPK_2		
I'm just trying to help the cities, see how they can. How they can collect data.		
[50:56] SPK_1		
Okay,		
[50:57] SPK_2		
because my company we don't own in the data. We don't collect a data. We just doing things for others.		
[51:03] SPK_1		

Okay.		
[51:04] SPK_2		
Yeah to use the data in a smart way.		
[51:09] SPK_1		
Yeah, yeah. Okay understand. Thanks and how about like people movements like citizens?		
that kind of social activity in something that I will that be considered as one of the data that using as well.		
[51:27] SPK_2		
It could be we don't we haven't come that far yet		
[51:32] SPK_1		
Okay		
[51:32] SPK_2		
It's kind of hard to store that data and I have a deal with the data at all, but I can think of is like the the GDPR like they yeah, how can be on personalized the data so we can store it? It's hard, but I know there there are municipalities in Sweden who have looked at traffic the movement from cellular phones smartphones throughout cities to see patterns. Let me Google that.	DC DSU	ED
[52:17] SPK_1		
I think it's <Location Information> that I read online.		
[52:21] SPK_2		
They could be that yeah.		
[52:22] SPK_1		
Okay. Yeah. Okay,		
[52:25] SPK_2		
and I know that the city that did it. They got like there was some sort of verdict that they couldn't do it anymore because it was some compliant with some sort of legislation. Yeah. So now we haven't come to that but the national way will be like how people are moving in the city and how traffic is moving today and then we come to how is water sewage but these these type of information is always restricted to use due to national security so you cannot like you you cannot have not even if you work within an organization, you don't have the possibility to look at all the sewages in city and you cannot like take that data and put it somewhere else or send it somewhere else. I know in <Location Information>. You can look at a area of 100 to 100 meters for like 15 minutes. Data is gone.	DSU DS	ED MD
[53:44] SPK_1		
Okay,		
[53:44] SPK_2		
it's such a from the national security side. If you're planning a terrorist attack and you know, where all the water supply in <location information> are the big lines for the water supply then you can do something evil. Hmm. Yeah. Says such things we need to also take in consideration. We're dealing with the digital twin.	DS	ED MD
[54:08] SPK_1		
Yeah, would that be similar issue with people as a people data and then or well		
[54:17] SPK_2		
I'm not an expert on this but just making a wild guess here. You can see with it with your own eyes where there are a lot of people.		
[54:27] SPK_1		
Yeah.		
[54:29] SPK_2		

That's not a secret.		
[54:30] SPK_1		
No		
[54:31] SPK_2		
But you cannot see where with your own life where the fresh water flow. Because that's on the ground so I don't think..could be a terrorist. That's a security issue with dealing with people data as well alongside with the personalized GDPR thing but I'm not expert on that. Sorry.	DC	ED MD
[54:55] SPK_1		
No. yeah, so we wanted to ask how groups and areas the groups of people in areas that are affected by the digital twin how they are represented. in the data, but you		
[55:14] SPK_2		
I can't answer that. Sorry.	BP BD	
[55:19] SPK_1		
No, but coming from the spatial data because it's like it's the data is already there. So, how can what do you have like kind of like guidelines and or tell you like how they collect the data so you can use it or is there something that they mentioned in the data? So, you know that what is the data distribution in a spatial data?		
[55:42] SPK_2		
Well, they there are guidelines how you collect data and if you collect the data, but actually or you collecting and how is this? How the quality of the data how you make how you communicate that that's already on the national level and there are guidelines for collecting these kind of information. But when that's when you collect data. The other thing is, where do you have collected data? Like let's take the tree example, which is quite a good you have. In <Location Information>in the city know where the trees are that are in the city center. But when you come out in the forest, there's no municipality to have all the trees in the forested forces either database. Where is Where's the the line between we have full control of where all the trees are and we have no idea where our trees are. that's not like a hard line where the thing one thing occur and the other don't and how can we in one way say How good is our data?	DC DQ BD	TCA
It's really hard. because perhaps that data of the trees are collected during the last 30 years and		
[57:13] SPK_1		
and then if we go back to the decision making around this, do you know like how much of the decisions reach out to those areas where less data is collected or is are there any consideration on How much decisions can be made? Despite maybe data.		
[57:34] SPK_2		
The thing is but when you see the need to do something in those outskirts of the municipality they go and make like an update through their data just for that area so they can make the decisions just for that area. But if you need to do something on from the whole municipality like how many trees does <Location Information> municipality have then speaking of taxes and how how things are costs. Doesn't happen.	DC BD	PD
[58:14] SPK_1		
OK.		
[58:17] SPK_2		
So go ahead <Name>.		
[58:20] SPK_1		

Do you have anything like?		
[58:23] SPK_2		
Oh, yeah, I can speak forever.		
[58:25] SPK_1		
Yeah. So how about like So if you see that the data quality is not as good as you expected. So was it decision that you want to make or like what kind of things that you will want to do and trying to like improve this situation? Sorry, is that hard question again?		
[58:54] SPK_2		
Well, it's hard question because but in the municipality you will always have like the decision to take is it worth it.	BD	PD
[59:01] SPK_1		
Hmm.		
[59:02] SPK_2		
Is it worth that? I take my colleagues? Go out in the middle of nowhere. Spent a week collecting data to make the database better. And then what which that we have like a small area where it's really good. How do we keep the data quality that really good there. Should we have the same approach in all areas. Can we afford that? How how high should be reached in data quality should it be really good certain places or and worse? Should it be on the same quality you throughout the whole city? I can't answer that	DQ DC BD	PD
[59:50] SPK_1		
that's again a question about like the best outcome for the least amount of money and yeah,		
[59:56] SPK_2		
which is kind of sad.		
[01:00:00] SPK_2		
It would be better if everything was perfect.		
[01:00:02] SPK_1		
Yes. Ok So money is definitely an limitation running these kind of projects		
[01:00:12] SPK_2		
Yep definitely		
[01:00:13] SPK_1		
should we go into? Yeah this a bit Yeah, so we wanted to ask what's your current understanding of algorithmic bias? Is that		
[01:00:25] SPK_2		
I have no idea what that is. So so no, please explain what actually you're looking for. algorithmic bias		
[01:00:36] SPK_1		
Yeah. so I guess that's That could be issues in quality of data like distributed differently towards different. People or areas which makes predictions more accurate for certain. Groups or areas and left . It could then how do you say that unproportionally effect? Or disproportionately affect some people worse like then there's some kind of bias in the data or in the algorithm or built in the the system Somehow		
[01:01:14] SPK_2		
Then I understand What it is. so if you take an example just to make sure if I understand if you if you don't have any information about bike Lanes. Then all your predictions of traffic will be towards cars.		
[01:01:28] SPK_1		

That's a very good example, I would say. Because to add on that is that because the data you collected from time to time. You will also use it to train the models and then to give out a prediction. So the bias may be built from the input data and then goes into the training models and then go to prediction as well. So that's what we think about algorithmic bias		
[01:01:51] SPK_2		
I can understand that and from my perspective when you're dealing with with spatial data citywide. We're not in that kind of detail at all. I will suggest speaking someone who is more like in traffic planning or more like in the trying to solve a problem or like trying to to see where should I build certain things. But of course we can see if I come take away the the digital twin side and look at <Project Information> where to place solar panels if you don't have the how the roofs are constructed if you just have boxes. Like square boxes with all the roofs are flat. Then you don't give a respective is. decision of where to place solar panels because every decision will be wrong because there's no houses with flat roofs and But I cannot say that we have found we're not there yet. That's further ahead. Like when we doing analysis and digital twin because there are really no digital twins in that detail level. yet.	DQ BD	ED MD
[01:03:07] SPK_1		
Okay. So like in the projects, this is not something that you are discussing like regularly or is you're not there with		
[01:03:17] SPK_2		
no. We're not there yet. Sorry, but it's a really good point. actually		ED MD
[01:03:24] SPK_1		
So actually I could have a follow-up question about that you mentioned about the national standard. So is there a industry standard that you got from dealing with data things?		
[01:03:37] SPK_2		
There are industries there is a aim to be in the industry standards in the spatial world or like all the data that are within. The planning of new societies constructing and planning of the society. All your spatial data that is they have found 150 different data sources of different types of data that they're trying to standalize. It could be everything from there like them the master plan for what you should you can build within a municipality to how shall we What's the national standard for? depth measuring within a lake or 3D scans from a from a helicopter over Forest house. You need a National Standard of how this information is built and the and yeah. But I can send you some information in the chat while we speaking? but but <Location Information>. Me as from the private side of things but also were <Name><Education Institution>	DQ DSU	
[01:05:29] SPK_1		
Okay,		
[01:05:29] SPK_2		
Have you met him?		
[01:05:31] SPK_1		
I don't recognize		
[01:05:33] SPK_2		
because he's really heavily involved in the like digital twins in Sweden and the situation of how to deal with national standards and such and I think it's really good to speak to and it's also quite friendly. Yeah.		

[01:05:50] SPK_1		
Maybe we can get a hold of him. Yeah, actually. Interesting so I was thinking about this since it's not like the data is not that detailed. So that bias is something you discuss that much.		
[01:06:16] SPK_2		
[Shaking head]	BD	TCA
[01:06:16] SPK_1		
Are there then any like how much? Is that a consideration when you make decisions like the detail of the data and then how detailed decisions you can make kind of if that makes sense?		
[01:06:29] SPK_2		
Yeah, definitely well. Well. all the time, I will say oh always well, the detail is detail expected the data quality expected to things but also how good is the the quality like geospatial wise is is there certain geographical areas where we don't have the best covering of the data so, of course, there are a lot of different decisions made from these kind of aspects, but I will say to get a better answer to that you should actually speak to an architectural or city planner or a traffic planner or someone who actually do then else is that will affect different decisions within the city	DQ	ED MD
[01:07:22] SPK_1		
now we have we have actually traffic and city player		
[01:07:26] SPK_2		
great.		
[01:07:26] SPK_1		
So hopefully that might lead to something then. And yeah, one thing that we can ask is you talked a bit about it already, but we were thinking about when you collect data for a certain system as you mentioned before and you try to transform it into another system. What Are the considerations there? Like how do you manage that kind of move of data?		
[01:07:59] SPK_2		
Yeah. Oh there's a lot of different aspects how and both geometry wise and attribute wise the the geometry of the data if it's in one way in one system and that geometry should be in another way in another system. Then you need to redefine the geometry in some way if you have like that's kind of hard because if you're done it if you have it like one like a building a 3D building and then you do something with the geometry fit in your solution. Then you need to work together with this redefine data you do things analysis on that. And that this building got an update.	DSU	
[01:08:55] SPK_1		
Hmm,		
[01:08:56] SPK_2		
how will that affect this building and the things that you've done with this building that that's all dead because here is a new building. so that's the one thing another thing is like attributes how if you have one thing in your source data like this building is for hotels and restaurants But in your where you want to store your data, you have two different. Like places one for hotels and one for restaurants and how do you split the data? Because it's just one building with two purposes.	DSU BP	
[01:09:40] SPK_1		
Yeah.		
[01:09:42] SPK_2		

a lot of these questions arise when dealing with like the schema mapping as we call it like defining the schema and change the schema from source to to the goal. But I could always be both for the attribute side and geometry side.	DSU	
[01:09:57] SPK_1		
Yeah. But so then going back to the bias. Do you think there might be a risk of like the people then transforming the data that they have certain like interpretations that could be built in that are not like what are the standards for?		
[01:10:18] SPK_2		
Definitely everything that's involved people are useless. Like people take decisions. That's what we are good at. Which means that If you and I do the same job, we would come up to the different decisions. Definitely that's a risk of dealing with data reconstructing data. If the data model is already information models isn't exactly alike. We cannot do like a one-to-one mapping. We need to do something to map from to and you need to deal with with a with a human mind or human brain or human humor because it could be one decision in the morning and another decision in the afternoon depends on how tired or so definitely	DSU BP	
[01:11:00] SPK_1		
yeah. Is there some some kind of processes to like this is Individual decision then or is there a process to come up with a consensus how you should transform the data?		
[01:11:15] SPK_2		
I will say that. you have a consensus decision in on the theory side, but then you start working with things use you realize that the theory isn't actually appliable in the real life situation. You need to take a lot of different decisions and then the the time and the budget is kind of narrow. So need to take fast decisions to move forward. And then they are people decisions made there.	DSU	PD TCA
[01:11:48] SPK_1		
Okay.		
[01:11:54] SPK_2		
But hopefully they are in some way and written down. And I don't know a feedback so worth where you could say that we took this decisions. But if we will have consensus on everything we do we won't get anything done. We need to move. We need to move and then take decisions and revise what we actually done	DQ	PD TCA ED MD
[01:12:27] SPK_1		
And I was thinking then if certain data is collected towards a certain goal from the beginning and then to another system that has another goal.		
[01:12:35] SPK_2		
Yeah,		
[01:12:36] SPK_1		
Do you think that can be an issue.		
[01:12:37] SPK_2		
Definitely.		
[01:12:38] SPK_1		
Yeah, how		
[01:12:39] SPK_2		
Well well that this is a good decision. A good example, take a building. Because a building is like the villa you have in your like the family houses in <Location Information>. They are always like rectangular and there are roof look like this and they are red and white, but if you have a roof like this if you say that the roof height is this level or this level where the roof reaches the house and you take that information put in a database and all everybody who knows that. Okay, this this	DQ	

data with a but if you put the data and put somewhere else in other context and you said just roof elevation, perhaps somebody thinks it's that elevation. Hmm. Which means that you take a decision based on? Wrong height of the roof and the most scary part is when you mix this information, you have one database roof height and some have this top elevation and some have this bottom elevation and isn't in the same database. They have no information what where on the roof you have to you have collected your data. Of course, they're gonna be a lot of different mistakes if you take decision on that data		
[01:14:09] SPK_1		
So that's a bit of a communication problem.		
[01:14:13] SPK_2		
It could be a communication problem, but it could also be a software problem because the software perhaps cannot put any metadata on the fields and say where have I measured the roof elevation. Now and one other thing is that now we have done one thing measuring roof elevation up here say in this municipality without work for and we've done the same thing for 30 years and then we find okay. We should also measure this elevation too because then we get two elevations that could be useful in some way. How then we're going to get a situation where you have all data collected for like 30 years and you have new data collected from now in the same database and and how can these data does live together or should we throw all the old data out? And then just add in the new data? And there's a lot of things of changing the way the municipality or the city work. They have to do what they have done for a long time parallel to a new way. They should do things and how can it some live in that ecosystem with a new and an old solution in parallel? Perhaps they have old like ongoing projects that need a data to be restructured in the old way. That you have projects that deal with the new data the new way and this project can go on for like 10 15 years. Yeah, then you can understand this kind of tricky. Sorry about I think I should yeah, it's almost 2 o clock and I think I should move along here.	DSU DQ	
[01:16:09] SPK_1		
Sorry that this took more time than we expected.		
[01:16:12] SPK_2		
I think it's my own fault.		
[01:16:15] SPK_1		
No. No, I was just wondering if we have any questions that we felt we didn't get answer to is it okay to email them to you?		
[01:16:27] SPK_2		
And of course? Yeah.		
[01:16:28] SPK_1		
Okay, that's good to know so we don't have more time. Yeah, I guess thank you so much for today. And your answer is very interesting.		
[01:16:35] SPK_2		
Oh hopefully Just to be clear. I'm not working in the city. I'm up from a private company helping the cities to do things that they want to achieve so I can see things from the outside.		
[01:16:51] SPK_1		
Yep.		
[01:16:54] SPK_2		
Yeah, that's good to know.		
[01:16:55]		

Yeah.		
[01:16:56] SPK_1		
So I think we called it. Yeah.		
[01:16:59] SPK_1		
Yeah, great. Thank you so much and have a nice day ahead a few remaining hours for your work. Yeah,		
[01:17:06] SPK_2		
yeah, too. Take care.		
[01:17:08] SPK_1		
Thank you. Bye.		

Appendix 4: Interview 2, P2

Theme	Colour Codes	Factor	Text Codes
Context	Yellow	/	/
Data Management	Blue	Data Collection	DC
		Data Storage & Use	DSU
		Data Sharing	DS
		Data Quality	DQ
Algorithmic Bias	Red	Data	BD
		People	BP
Principle of Discourse - Democracy Principle	/	Pragmatic Domain	PD
		Ethical Domain	ED
		Moral Domain	MD
		The Theory of Communicative Action	TCA

Participants	
Authors	SPK_1
P2	SPK_2

Transcript	2nd layer	3rd layer
00:00:01 SPK_1		
OK.		
00:00:04 SPK_2		
I was just trying to see if I can change the video but it doesn't seem like that.		
00:00:16 SPK_1		
Oh, is it on the right corner there is a view, then you can change the window.		
00:00:22 SPK_2		
Yeah, I don't seem to have that function or working right now. I have. I'm only the laptop video.		

00:00:35 SPK_1		
Okay, okay, okay. So starting with some easier questions, do you want to tell us a bit about your role right now, what you're doing from day-to-day work?		
00:00:55 SPK_2		
And as I have tried to describe to my children, I sit in front of a computer, I serve the web, I answer and send emails and I sit in meetings. But like more professional description of it would be that I'm a <Role> I help clients basically with three different things, purchasing IT architecture or project management and that is always around geographic information or IoT. So that is I would say that is my role. So I have I helped various clients within those two fields with those tasks.		
00:02:01 SPK_1		
OK, so do you want to tell us a bit more about the Digital Twins and specific the projects surrounding Digital Twins in the city setting?		
00:02:13 SPK_2		
And I've been I have had basically two projects that I've been focused around Digital Twins. One of them was for <Company> or <Company> ==where we had a workshop series on the usage== or the well, ==why digital twin would be a viable solution for cities in order to support the decision making and then getting further with their processes.== So that is one of the projects and the other project is working on. On that project is ongoing is working with the municipality of <Location Information> and helping them to take the first steps towards digital twins.		
00:03:29 SPK_1		
OK. So in these projects, what are the organizations and companies or people participating? Except for you and the municipality then.		
00:03:43 SPK_2		
Well, in the <Project Information>, it was, I think we had around <Number> participants from different municipalities and regions and of course there was some like ==system providers or companies== like that. Involved and in <Location Information>. It's only so far people from <Location Information> and there we are looking at like different proof of concepts for digital twins.		
00:04:21 SPK_1		
So yeah, I'm sorry. You continue.		
00:04:23 SPK_2		
the first question that always comes up is ==what the hell is the digital twin?== And from my perspective it's there is no good answer to that question. Well, I have not got from anywhere a good answer because it it's dependent on your perspective if you talk to somebody who works with the Urban planning, they want to use a 3D model or visualizing what is, what the plan is going to be and for them that is the digital		

twin they have like a thread model of the houses and maybe clone them with some realistic they take the BIM models and put them together in a in a mesh and yeah. Whoops, the digital twin of the future or something that is not there. And then you talk to somebody who is working with real estate and maybe running couple of couple or a couple or hundreds or thousands or buildings and they want to be able to use. Digital twin to see how the energy usage is or how to follow the things in real time. And those are two totally different use cases with totally different requirements for the base of what the digital twin is. And then you go to the other extreme and go to a car producer that has made a digital twin of a fossil engine and so that they can model exactly what happens with the fuel inside of this engine while they are designing it. A totally different use case.		
00:06:52 SPK_1		
So even though that there is different perception or the definition about Digital twins, So what are the overarching goals of those projects because you mentioned once is about supporting decision making, so is there other overarching goals of those digital twin projects?		
00:07:12 SPK_2		
Right now, the overarching goal is a little bit. We don't know what the hell this is. We need to learn. Because I don't think there is a single municipality in Sweden today that has a digital twin solution that is for general purpose.		
00:07:51 SPK_1		
So it's more for specific cases.		
00:07:53 SPK_2		
Because they are like more specific, they are more for, yeah they have like created a digital twin for some, some have made like for certain areas of the municipality of the city very detailed like on BIM level detail and some have made like for the whole for the biggest Urban areas in the in the municipality, mostly with the thread models, like when you're talking about LOD, maybe two or something like that. I think LOD2 is a thread models, right? Or is it three? I don't remember. It depends on which standard you go to and I think one of the municipalities that has come further with this is <Location Information> and you probably have <Location Information> on your list. I assume <Name>, they in my opinion they have focused a lot on like showing it that it is. That it is for like they have put the data into <Platform> so that you can get really cool visualizations. But in the in the <Company> Workshop series we found that there are that the digital twin has to support. I think it was four main use cases. One is visualization. Another one is analysis to do, for example urban urban heat islands analysis or Skyfall analysis. That comes a lot of rain or that kind of wind analysis and things like that. All those kind of analysis that you can do. The third one is that it is a part of of the unbroken digital chain for urban planning. Because in order to have a unbroken chain that is like when you make the detailed city plan or over over city plan in Swedish. Don't know the words of those in English.	DC DQ DSU	
00:11:00 SPK_1		

Never mind, she will translate it.		
00:11:05 SPK_2		
When you have those, you want to have like an unbroken digital chain so that you can go from a detailed plan to a complete building permit and accepted house so that that is an unbroken digital chain. Now this chain is through a lot of PDF's and like going back and forth from another digital documentation. So that is that was the third use case. The 4th use case is for real estate and actually the 5th use case is for IoT that you want to use real time IOT data to to do different things like to monitor your city what what is the state of things in general. and all of those different use cases, they require different things from the digital twin. Yeah. Was that an answer to your question?	DC	ED MD PD
00:12:21 SPK_1		
Thanks. Yes. That was answered to a few questions, I think. Yeah. So that's why we are reading the question and trying to skip some questions. Sorry, this is our second one. That's why I'm trying to get used to this format. Yeah. How would you say that you decide on which areas to implement the digital twin, like which geographical areas do you? Is there some kind of discussion for where it's needed? And so on?		
00:13:01 SPK_2		
I think right now I have not heard anyone claim that like a digital twin is absolute is a requirement for doing things that you need to do, the requirement today is more that you need the geographic data to do the things that you need to do. The promise that the digital twin brings is that it offers you all the the geographic data to do those things. I think that is. That would be a more that would be my perspective on the question. So and as I said from what I see most of the organization that are discussing digital twins they ask because digital twin has a lot of like questions to solve and the digital twin is very undefined in what you need to do. Like for example you have the level of detail, the LoD. Are you going to use LoD three or four or two or your digital twin? And that is only one of the things that in my perspective one of the most important issues to solve. Is the difference between the geometric digital twin and the semantic digital twin. Are you aware of those differences?	DQ	
00:14:58 SPK_1		
You if you can explain it, that would be perfect briefly.		
00:15:04 SPK_2		
The geometric twin is like the shapes of things, that it's the geometries and the the the. The shadings, the colorings and the pictures on the houses and all of that. The semantic twin, that is how you describe what an object object is and how you get further information. The linkage to other information sources, for example a digital twin would it would not be very wise. To keep the state of a light bulb of a light post in the city in digital twin if the light post is on or off. But you would fetch the state from the light post system where you have data system that that keeps track of the light post. They maybe know that yeah, at this time of day. We keep it at a dimming level of 50% or that this is a light bulb that is only on and off with this lumen	DC DSU DQ DS	PD

strength or something like that. And that is a lot of information that is relevant for the or everybody who's using the digital twin, but the link to the the specific move and the information about that is important. In order to keep the digital twin intact, did you? So that is the semantic information. Where is the light bulb? It's like the geometric information, and all the information that is needed to get the state of the light bulb is the semantic information. Addresses can be semantic information, number like the identification of of different apartments, identifications of light posts that may be needed to do some of the analysis if you want to do or visualizations or whatever. And here you come to one of the things that is. Or a municipality, A major issue, which is how do I get this information and how do I keep it accurate all the time? Because I mean keeping the accurate information of the exact form of the geometry of the light post is one thing, but keeping the links and Information meta meta data about the bulb so that you actually can use the information. It's another issue where most municipalities don't have neither technical structures nor organizational structures to handle data sources that are need to be updated in different parts of their organization because most of the time it would be the geographic department that would be updating the geometries and things like that. Whereas the light post information the the on and off state and metadata about that is from the traffic department. So you get my difference here.		
00:19:06 SPK_1		
Yeah. So is there you said that this is an issue for the municipalities. Do you know this is one of?		
00:19:13 SPK_2		
This is the main issues that municipalities are not gonna have when it comes to building digital twin in a big scale.		
00:19:27 SPK_1		
So do you know if there's anything that they're working like specifically to prevent this Issue? Is there a solution to this?		
00:19:38 SPK_2		
I I don't know the. I haven't talked with <Name> about their experiences of this and how how they are working on it. The approach we are trying to use in in <Location Information>. is to make smaller proofs of concept. Where we take a little subset of informations, say for example light posts, where we are looking possibility of taking the light posts and making proofs of concept that we can take a state and separate those the geometric and the semantic handling and link to different system and see how that works and to prepare the organization for handling that find the best or the most suitable organization and solutions for that.		
00:20:41 SPK_1		
OK, so yeah, OK, so it's more like you guys are trying to come with some new technical solution or proof of concept and trying to solve this kind of conflicts or issues.		

00:20:54 SPK_2		
Well, I don't think it's not a technical problem. It's more of hmm it's more of an organizational issue. Different organ parts of the organizations will be have to have to cooperate with maintaining the data, keeping the data maintained. If you go to our <Platform> or some other systems provided this is a no brainer, it's just just by our system do it. But for for the municipalities this is the issue. Here is about how you build and maintain information of such detail that. You that you can keep it. That you can keep the accuracy and actuality of the information over time, because as soon as the information is not updated and not actual, people stop trusting it, the users won't trust it. Yeah, yeah, we have this digital twin, but I know that I can use it and it's not Digital data. I tried it. OK, then it's Then it's you first. Yeah. Then it's throwing away your money.	DSU DS DQ	PD
00:22:27 SPK_1		
OK. So decisions that are being made based on the digital twin, it's like a consideration with the data quality, how much decisions you can make based on the digital twin.		
00:22:38 SPK_2		
Yeah. There you compose to. The issue of level of detail are you, I don't know your what education you have or what kind of courses.		
00:23:04 SPK_1		
we are in information systems informatics?		
00:23:07 SPK_2		
OK. So when I say level of detail, what does that tell you?		
00:23:12 SPK_1		
I will be thinking about the data granularity for me. I'm thinking about that.		
00:23:18 SPK_2		
And that's data granularity is a good thing. So when we are talking about for example for digital twin, if you have a geometric representation or a digital geometric representation of the city if you want to do skyfall analysis depending on the slope you're working with, the more slope, the lesser accuracy you need, because then all the water obviously flows down the slope. But as the ground becomes flatter, as most cities are, the more detail you need to have on your data in order to know where the water is going to flow or when it's going to stop. And in order to do a proper such analysis that is worth something and you need to know exactly the dimensions of the pipes that are underground, that lead away in the water. And you have to have in the worst case, you have to know the heights in millimeters or at least 70 meters and that gives you a detailed level or data granularity that is extreme. And when you're working with that kind of application, and if you have too much granularity, that means more data, heavier processing	DSU DQ	PD

times, more difficult to process the data. If you're uninterested in all that level of detail, then that is just a burden for you. And the same thing for. Like if you are working with urban planning, the level, the detail level of of the ground is not so very interesting. But there you have it's a lot more exciting and interesting to have high level of details of the facade of the buildings, which is totally uninteresting for the one who is doing the skyfall analysis. So the question then becomes, where do you find the least denominator that fits all? But I don't think anybody has come to a proper answer on that issue.		
00:26:01 SPK_1		
OK. So if we relate that back to the goals of using the digital twin, how do you reach a consensus for which goals to prioritize? When you need to do different things with the data to reach different goals, if that makes sense.		
00:26:20 SPK_2		
Yeah. And I don't know.		
00:26:22 SPK_1		
OK. There's no process or anything to handle that right now or how is it worked with?		
00:26:29 SPK_2		
No I think if it comes back, it comes back to your use case. What are the? What are the? Prerequisites that your use case needs. And in order to do the Skyfall analysis, you don't need to so that you get a proper result and the result is dependent on other. As mentioned, it can depend on the slope, it can depend on what kind of how, how secure. Or yeah, how secure do you want your analysis to be? Is it okay that it is only 80% right, or do you need it 99% right? There's a huge difference in the level of detail of data you need for your of those scenarios. Does it come? It goes back to what the specialists that do the Skyfall analysis need for data. They will need data about what kind of soil it is. Is it sand or is it mud? Is it stone? Is it is it asphalt? Or is it a park with grass? Because the water case behaves differently in all of those kinds of factors.	DQ DC DSU	TCA PD
00:28:01 SPK_1		
Yeah. So, sorry		
00:28:03 SPK_2		
Yeah, what type of scenarios? What… Yeah, all of that.		
00:28:09 SPK_1		
Yeah. So who is included then in making those decisions? You mentioned that specialists, who else is included in discussing these kind of things?		
00:28:30 SPK_2		

I guess that when you come into those discussions you. You will have the specialists on that specific topic, for example skyfall or urban islands or or whatever. And then you will have the geographic department involved because they have already a lot of the data but not all of it. And then you have decision makers that make a decision. Okay, we are not ready to pay 500 million for this analysis we are ready to pay 100,000. So I think those are the main group of staff would be involved in that.		
00:29:15 SPK_1		
So you mean Decision Maker is the city planner?		
00:29:19 SPK_2		
No, it's usually their boss.		
00:29:27 SPK_1		
So what do you think is important when designing? Or implementing a system that is used for city planning? Personal opinion? Yeah, I know it's pretty broad question.		
00:29:41 SPK_2		
Can you Take it out again		
00:29:42 SPK_1		
What's the most important thing to consider when designing or implementing a digital twin in city planning purposes?		
00:30:11 SPK_2		
I think I'll pass.		
00:30:14 SPK_1		
OK. OK. That's a big question. Yeah, I.		
00:30:18 SPK_2		
I think my answer would be too much of the guess guess.		
00:30:22 SPK_1		
OK.		
00:30:22 SPK_2		
For that, yeah, I can. I can give like what I would think it was, but I don't. It's not necessarily what a city planner would give as an answer.		
00:30:40 SPK_1		

But it's it's interesting to see your perspective of it as well.		
00:30:54 SPK_2		
And I guess it depends on if you are what kind of city planner you are, Are you working on the local detailed plans or are you working with the Overall plans. Are you working more as, say, a city planner for traffic, or are you working more as a city planner for parks? But I think that today they are not using digital twins a lot. On this then what they might be doing is that they might be doing using 3D models of with threads or with the buildings for this. I think that is basically where they come. So then again I think many want to have a digital twin where you can do more.		
00:32:00 SPK_1		
So they have different goals depending on which perspective or which role you have.		
00:32:05 SPK_2		
Yeah, I think so. Depending on on what kind of planning you're doing.		
00:32:10 SPK_1		
Yeah, Do you know?		
00:32:12 SPK_2		
Somebody who is working with traffic planning, well, the traffic flow is is important to and to see what is the usage of parking lots if you're working. With parks, and green areas then that kind of areas and how many people, what is the, how far do you have or the average add to a green area bigger than one hectare or something like that?		
00:32:56 SPK_1		
Yeah.		
00:32:57 SPK_2		
I think it's really dependent on what your subject is.		
00:33:02 SPK_1		
And you mentioned before that there's. It's not really a way to decide which goals are more important, right? Or is there a process for that?		
00:33:13 SPK_2		
I am not seeing the process for that anymore. Okay that anybody has. No, I'm not seeing that. I think most of those analysis that are done today. Now they are mostly done with like geographic data available through different sources, they are done by specialists within those different fields.		

00:33:47 SPK_1		
Okay. So I think I heard about that you also talked to <Company> before about <Company>. Is it <Company>? OK, maybe I misheard. Oh sorry, because I'm wondering so when you are implementing the technical solution, is there any involving other for party like contractors instead of building that in house in <Company>only?		
00:34:18 SPK_2		
In the project where I've been involved that we have not built any digital twin yet.		
00:34:23 SPK_1		
OK, so just proof of concept things. OK, understood. Okay. And should we go into data? Yeah, Okay. So we wanted to ask you what kind of data is being collected? We kind of touched on that already, but. What kind of data is this relevant? Did we already? Yeah, what we heard about is more on the spatial data. So is there other data that that will be collected to like complementing the solution in digital twins or the proof of concept that you're working on?		
00:35:10 SPK_2		
Yeah, we are looking at in <Location Information>. We are looking at how can we combine different data sources, so that we create and. So that we have those links between the data sets or the systems or whatever you like to call it, but so that there are links that an application can use in order to like get the real time data or get more data about a specific feature in the digital twin.	DC DSU DS	
00:35:47 SPK_1		
So it's not only about the data that already is collected by I would say other department or something. So it's also have real time data as well.		
00:35:59 SPK_2		
Well I mean you can yeah it can be real time data for example if you would like to have. Like sometime in the future you might want you are the boss of the municipality and you want to see everything that happens in the municipality in real time you would have, you would be able to see the buses running on the roads in a like unreal world and and for that you would have to know what the yeah, the color of the bus, the size of the bus and things like that and that isn't like not a type of object that will be moving around in your world or in your digital twin.		
00:36:47 SPK_1		
Okay. Is there a plan to do the same thing with people, people's movement, or is it mostly traffic?		
00:37:00 SPK_2		
I think there is an interest on doing that with people's movement. What I have seen is that, for example, planners of green areas are interested in seeing motion data for	DC	MD

parks, so that you can see for example that there are to this park for this park area. There are coming 100 persons a day or 500 persons a day or and that is important for both for how you plan your maintenance for that park, but also for how you should plan or alter the some other park that you need to. That is somewhere else, some somewhere else in the municipality or in the area. So movement data people is equally important. I think of bicycles or planning bicycle roads or you look at that for for example for safety where you should plan, for example, better light posts or some things like that to to increase the what we call upplevt perceived safety.	DSU	
00:38:36 SPK_1		
So for collecting this type of data, is there any considerations of how well it's being represented in the digital twin? Or that's maybe speculative. I'm thinking if we collect data about people and movement and yeah, people and areas, how well they are represented in the data. Or how some I don't know how to put to put it. Do you understand what I mean? I just for example like how for in the area of the data maybe here is more than the other areas of data and then how good is it representing the real life?		
00:39:28 SPK_2		
That's like if you have a If I understand your questions right, I try to explain how like if you have a motion sensor like gives a tick on off on off for when people are there or when it tracks people. That kind of information is kind of useless to see on the map, like lots of ticks, tick, tick, tick, tick. So you would maybe like to do heat map or something like out of that and. Of course that is important because that is then you do some kind of analysis of the real time data and I don't think there is any like consensus on how you should represent that data in order to be relevant, because for one planner it's more interesting to see I I want to see how it looks. On Sunday mornings between 10:00 and 11:00, and for another one, no, I want to see the Because this is a nightclub area, Sunday mornings at 10:50 is uninteresting. I wanted to see it on Saturday nights at Saturday night and Friday night. That's most important.	DC DSU DQ	PD TCA
00:40:40 SPK_1		
OK.		
00:40:41 SPK_2		
So it depends on what you're looking at I think. Therefore you need to look at like. The least denominator of data, the most detailed level of data and then you do some kind of aggregations or analysis on that that you add to your digital twin.	DQ	PD
00:41:02 SPK_1		
Okay.		
00:41:04 SPK_2		
Was that a reasonable answer?		

00:41:06 SPK_1		
Yeah so I've got a question. Like when you have the data on hand, is there any like explanation on how they collect the data and then what is the metadata of each data attribute and something like that so you can more accurately to understand what's the data for or how it's collected. Is there any information about those data?		
00:41:32 SPK_2		
That is, what you come to know is one of the bad conscious of most data providers because they are not used to create proper metadata and the metadata and as as you say, you need to know the quality and not only the quality, you need to know how the data is gathered, how this is collected, It's actuality it's it's depending on what kind of data it is. Like for example if you have parking sensors. That sense if there's a car or not, you need to know what is the error in this? Is it like 1% of the cars that park that don't get registrated or is it 20%? That means a lot, quite a lot when you are making decisions with that dataset because. As soon as you start doing analysis, the first thing you learn is that you have to know the quality of your data shit in, shit out. Otherwise or it will be and so yeah, that is metadata and that is also one of the issues that that many have problem with. And like the with many sensor gathered data, you get the the accuracy. Often only can be described on data quality, only can be described on sensor level. So if you have say that you have 5000 sensors that are sensing motion in in a city, in parks all over the city and half of them or 1/10 of them have an error of five percent, 1/10 of them have an error of 10% one, yeah and so on. How do you describe that? So that the one who is doing analysis with this data can get a confidence in that. I've not seen any proper ways of doing that.	DQ DC DSU	ED MD PD
00:44:20 SPK_1		
OK. So it means that there is no like industry standard that can be like used on like evaluating the data quality or?		
00:44:30 SPK_2		
I'm not seeing it. There are lots of standards, but I'm not seeing it.	DQ	
00:44:34 SPK_1		
Okay, yeah. So related to this, I think we can ask, what's your current understanding of algorithmic bias do you?		
00:44:57 SPK_2		
Are you talking about algorithmic bias as in AI? Taking biased decisions.		
00:45:06 SPK_1		
Yes or yeah, like processes with algorithms that maybe the data is somehow skewed so that yeah, the algorithm produces biased data or the algorithm in itself reinforces bias in data. Or the yeah, there's a few different kinds of the outcome is somehow		

biased towards one area or one group of people and makes better predictions for that part of the data than another part.		
00:45:48 SPK_2		
I've not come across that very much with geographic data. I'm not saying that it doesn't exist. But, but of course there is a as I was describing that you have those that you have different that's not I don't think that would be like classified as biased data what I was describing before or biased algorithms. No, I'm I'm not perceived that so far or as a big problem and I'm not like heard of that as a big problem. Then what I have heard, what I've seen and heard more as as problematic is the is that inexperienced specialists are using data that doesn't or they are they are not understanding fully or the quality of the data that they are using for an analysis and not understanding what affect that loss of quality has on their final analysis. From what I have seen, that is more of an issue that you have data that doesn't, but that you look at data as the truth and don't realize the effects of quality issues in that data following the outcome of your analysis.	BD DSU DQ	MD ED
00:47:57 SPK_1		
Do you think that's a communication problem or with this metadata issue that you talked about earlier, or what do you think that's dependent on?		
00:48:08 SPK_2		
I think it has to do with both communication issues but metadata issues, but also about that. Maybe people are not experienced enough with what you're doing.	DSU	
00:48:33 SPK_1		
Yeah.		
00:48:33 SPK_2		
Yeah. So I think it can be a number of factors.		
00:48:40 SPK_1		
Do you know if there's something that's being done to limit this issue or work against it?		
00:48:53 SPK_2		
Yeah, I think many, I think many decision makers, they want to understand and what are the insecurities or inaccuracies or risk levels that you are handling in your analysis. So that so that they become aware of that. Yeah, but I have not seen any like methodologies for that. But it is possible that those are taught in school today. It was not taught when I was in school.		TCA PD
00:49:34 SPK_1		

Okay, okay. And actually it's my question on because I heard about that you talk about IoT things the one of the use cases IoT. So is it using sensors to collect data or something else that is using as for the IoT use case?		
00:50:04 SPK_2		
Yeah, we are looking at those kind of proofs of concepts in <Location Information>. We are looking at how we can, for example, <Project Information> to try to establish a linkage between datasets of different organizations so that they start to learn how to work with that, how to maintain that kind of data.	DSU DQ	
00:50:42 SPK_1		
So actually I don't know is it because it's not really related to people, so doesn't really require any consent to collect data like this kind of things. Is there any things that we need to have consent for if we are doing the sensor or to collect data		
00:51:02 SPK_2		
To consent from whom?		
00:51:07 SPK_2		
Yeah, so that's why I think that yeah, maybe it's doesn't require consents for collecting data that is not related to people I think.		
00:51:12 SPK_2		
Yeah, I mean, always when you come to like if you're using cameras or things like that, then you come to the GDPR personal, yeah, yeah, GDPR.		
00:51:27 SPK_1		
Okay. So the parking data and all things, all the things that you mentioned, so it's not really related to people.		
00:51:35 SPK_2		
No.		
00:51:35 SPK_1		
It's not really, Yeah. Understand Okay.		
00:51:40 SPK_2		
And even if you measure like movements in a park, it's not related to people. They measure people, but totally unidentified.	DC	ED MD
00:51:50 SPK_1		

OK. Yeah, that's my question. Yeah, OK, understand.		
00:51:54 SPK_2		
So that is an issue for all municipalities that they they really don't want to go in getting in the information that you can where you can identify a person.	DC	PD
00:52:06 SPK_1		
OK.		
00:52:10 SPK_2		
Because then you have problems.		
00:52:12 SPK_1		
Understand, Okay.		
00:52:17 SPK_2		
And I'm sorry, but I have another meeting in 4 minutes. Are we?		
00:52:23 SPK_1		
Yes, I was.		
00:52:24 SPK_2		
Have I talked too much?		
00:52:26 SPK_1		
No, no, no. We didn't really know how much questions we would have time for beforehand. So this is a good measurement. I was just wondering if we have any questions that we feel that we didn't get answered cause we didn't have time to ask all of them. Is it okay to e-mail you. Yeah, okay.		
00:52:36 SPK_2		
Yeah. It's okay		
00:52:49 SPK_1		
Thank you so much. Then we might do that. Yeah, but okay. Thank you. So I think I've.		
00:52:53 SPK_2		
Got some? Good information from this.		
00:52:57 SPK_1		

Definitely.		
00:52:58 SPK_2		
And I am forward to see your piece of work.		
00:53:02 SPK_1		
Yeah, Yeah. Thank you for participating. Yeah. So thank you. Have a nice day ahead. Thank you. Bye, bye.		

Appendix 5: Interview 3, P3

Theme	Colour Codes	Factor	Text Codes
Context	Yellow	/	/
Data Management	Blue	Data Collection	DC
		Data Storage & Use	DSU
		Data Sharing	DS
		Data Quality	DQ
Algorithmic Bias	Red	Data	BD
		People	BP
Principle of Discourse - Democracy Principle	/	Pragmatic Domain	PD
		Ethical Domain	ED
		Moral Domain	MD
		The Theory of Communicative Action	TCA

Participants	
Authors	SPK_1
P3	SPK_2

Transcript	2nd layer	3rd layer
00:00:02 SPK_1 Okay. Is it traffic? Is it a traffic? Digital twins.		
SPK_2 Yeah. What it shows, the moving dots are vehicles. And the lines that goes out from them are packages that has a destination and goes to the destination. So what you can see is the trucks going and they drop off and they leave white spots as they drop things off and so the lines will be fewer and fewer as they unload their packages and so this is actually a digital twin to Simulate deliveries of packages at this time, That's what it is. And the reason to do that because talking about biases, we believe that there is a smarter way to do it.		
00:01:14 SPK_1 So you're trying to optimize the.		

00:01:16 SPK_2 Yeah, trying to show. What the gain is from doing it in a <u>smarter way and to simulate</u> <u>different kinds of solutions.</u> That's the goal with this digital tool.		
00:01:32 SPK_1 Yeah. So you're trying to communicate to someone.		
00:01:36 SPK_2 We're trying to <u>communicate to pretty much all the stakeholders in the logistic</u> <u>chain to gathered together in trying to do things in a better way.</u> The problem we have with this not the problem, but <u>this is the start.</u> So this is based on <u>statistic data,</u> <u>not actual data. Unless in a small portion it's actual data</u> because we got some data from <Company> and from <Company>, for instance, But they don't really know how their packages are delivered because they hire <Company> or <Company> or somebody else to do it. <u>So we have the origin and the final destination at this time,</u> <u>but we don't really know. So we need to put more data in it to be able to actually do</u> <u>some really good,</u>	DC DS DQ	
00:02:46 SPK_1 good digital twins. I would say,		
00:02:49 SPK_2 yeah, yeah, because <u>we want to do simulations on different kinds of solutions. But</u> it's a process because <u>it takes a lot of money and it's time and it takes the other</u> <u>stakeholders to want to join in with their data</u> and but. We have to start somewhere and we have to have something to show them so that <u>they will be willing to give</u> <u>their data.</u> Yeah, if you take the pointer, let's see if we do like this.	DS	PD
00:03:23 SPK_1 So to what extent do you feel like in the future they will be using real time data to process this?		
00:03:34 SPK_2 <u>Maybe real time data, but that's quite far away.</u> I think what you would do is <u>try to</u> <u>gather data for a certain period of time to be able to simulate.</u> But you see that you see the vehicle, it says <Company>, but it doesn't have to say that. It could say blue or red or whatever because it's not really. Because I think like <Company>, a problem with <Company> is if you talk to them, I talk to them and they would say sure, we'll help you with the data. That's what they say here. And then they ask the boss in <Location Information> and they will get a no. Because they're afraid that the boss in <Location Information> will say no. You know it's complicated. That's why we need something to show this is harmless. You. You won't. <u>You don't take a</u> <u>business risk by giving us this data.</u> So so <u>you don't have to write. Identify the</u> <u>vehicles. You can keep that secretly.</u> They can call them blue, but you can see. That all these packages are going and if we gather them at one point, this would be a smarter solution and try to simulate those solutions. So that's the kind of the goal of this digital thing that I work with. So I just wanted to show you, So we know what we are talking about,	DC DS DSU	ED MD

00:05:16 SPK_1 yeah, it's pretty cool because that's the first time we see the simulation in real life. Yeah, Yeah. The lines are getting fewer.		
00:05:24 SPK_2 Yeah, because they have delivered their packages. So it's actually on this data.		
00:05:27 SPK_1 So this is a simulation.		
00:05:31 SPK_2 Yes, it's a simulation. So if it starts over, you will see the same, exactly the same movements again. It's a 24 hour simulation. But it would be really cool to integrate the data to get live data all the time and integrate it with traffic measurings and maybe cameras and then identify the vehicles to show. Okay, this vehicle, it's driven by diesel and this is electrical. What are the environmental impact on this exact transport and possibilities are many, but there's a lot of work.		
00:06:23 SPK_1 Yeah. Because I saw that there's one data talking about CO2 emission. So I think that one is pretty nice and pretty cool to capture those data there.		
00:06:30 SPK_2 Yes. And our \<Department> they are looking into if it's possible to install cameras that can identify the vehicles. Not owners but vehicle types to see what are the emissions and yeah it would be kind of cool.	DC	ED
00:07:00 SPK_1 Definitely. Yeah. I'm thinking if we should..		
00:07:07 SPK_2 You know I there's you have a schedule hehe		
00:07:10 SPK_1 We have a it's a semi structured interview so we can jump a bit yeah. Yeah. But I was thinking if we should Go back to your role just for a minute, see just if you wanna tell us a bit about what you're working with, from day-to-day, what's the position?		
00:07:28 SPK_2 Position is \<Role>. For the municipality and the municipality do this because we want it to be a good place, to want it to be a good place to live and work and we need to have good conditions for companies to live because they create the jobs and they need a job to have a good quality of life. So our core business is quality of life really as a municipality and. But then there's this little tiny detail that quality of life is different for different people. So we have to balance around that. But jobs, everybody needs jobs. So we want it to be a good place for companies to live in. So we try to do that by giving, by trying to create a good place to have the company, in different ways, we have some part of that is talking to our colleagues in authority, parts of the organizations where you get bygglov.		

00:09:09 SPK_1 Building permits		
00:09:10 SPK_2 Building permits and things like that. So they do that in a company friendly way. Not that they should, but service minded. Yeah, and. One part is to try to create networks and connections between people and companies because we think if people talk to each other, development can progress.		
00:09:40 SPK_1 Yeah, yeah. Cuz I was wondering who is like participating in these discussions? Is it the companies and the municipality or who's the stakeholders?		
00:09:54 SPK_2 Stakeholders are Everybody, everybody. No, we try to for instance, we have something called the <Project Information> where we try to gather companies and organizations from the entire logistics chain or supply chain all the way from. I don't know if there's equivalent English expressions från jord till bord, från bord till jord, to make this circle away.		
00:10:33 SPK_1 Yeah. Yeah, you can translate it later. Yeah, it's fine.		
00:10_37 SPK_2 Try to explain that.		
00:10:40 SPK_1 No, actually I browsed the <Project Information> website, so I actually know about that. Yeah. So it's fine.		
00:10:47 SPK_2 And we do that. So the stakeholders are really, from the farmer, if you take food for instance as a product, from the farmer to the producer to the sales and everything in between and back to circular use. So and you can accomplish things if people, if we can just make small, because the supply chain is a lot of different stakeholders connected in the chain and if this first one can make some effort to make things in a more climate friendly way for this customer or her customer and this person can say, hey you can help me in every transition, can make just a small change, a lot of things will have happened at the end. So we tried to make these people talk to each other to see what can we do. I talk about <Company> again because I recently visited them. So that's why I have fresh examples. But they have, for example, they deliver the goods for <Company>. And they have very strict times where they are supposed to be delivered so like 10:45 sharp, and I understand why because then they can put in their schedule that somebody can be there and so, but <Company> has a lot of delivery spots and if they wouldn't be as strict on time, they could Instead of driving with five vehicles half empty, they could drive with two or three full because they can plan the deliveries in a different way. But it's not really an issue because they do what the customer asked them to do. But if the customer knew. That the price of this in a CO_2 way is this much and you could actually save this much CO_2 by just..		

00:13:22 SPK_1 Adapting a bit		
00:13:23 SPK_2 By adapting a little bit, maybe not if it's, I mean if it's a production chain, it's <Company> and they need the deliveries on exact time. But sometimes maybe that's not the case. And you can revalue because we have a new currency in the business and that's CO2, And that currency is the exchange rate is growing for the CO2 currency.		
00:13:52 SPK_1 Yeah, that's an interesting way of putting it. Yeah. So you how about citizens then Cuz this is about companies and connecting them together.		
00:14:03 SPK_2 Yes, So the department I work in is very much focused on businesses, Okay, and trying to make businesses run. So it creates jobs, yeah. So that's our main focus. And also visitors, because the visitors are generating money for the businesses.		
00:14:30 SPK_1 So tourism?		
00:14:24 SPK_2 Yes, But the city overall, of course, the main goal is for the citizens. Yeah.		
00:14:43 SPK_1 And citizens are in companies as well.		
00:14:47 SPK_2 yeah, companies consist of people.		
00:14:52 SPK_1 Yeah, exactly. What was I thinking? Yes. If you then have different goals for different companies, how do you, what's the process for reaching a consensus among those goals? Like is there? Is there a discussion or is there any weighting for which goals is higher priority than the others?		
00:15:12 SPK_2 We can't really decide for the companies. No, we can't or we cannot. We can't really decide very much. We can make some decisions we're looking into can we decide to make The center of <Location Information> for So no fossil vehicles are allowed for transportation, for example. That's something we could do. But then that's the same for everybody. Yeah, Yeah.		
00:15:56 SPK_1 And that's with policy?		
00:15:57 SPK_2 That's with policy. Yeah. So that's things we can do, but we can't really tell this company to work with this company. They have to find that. They have to find each		

other by themselves. But we can create a platform where they actually meet and we can insert knowledge and inspiration to do things and just hope the best		
00:16:29 SPK_1 Yeah okay. Hope that they communicate and		
00:16:30 SPK_2 yeah yeah networking and to because they, my, it's a big difference now from Comparing to five years ago when it comes to sustainability issues they have really like, just we need to do things. So they are very open to try different things and so and I think that's why <Project Information> is a success in that way that, We got a lot of companies participating and they and they really the only thing they really sign when they sign into the <Project Information> is an open mind to cooperation and to engage into the meetings with people from a strategic decision level, yes. We're trying to actually get the CEOs to meet. it's not always possible, but on a higher level because when they meet, things can actually happen. If it's just a logistic manager meeting, they can kind of sort of have an idea about things, but they can't really decide and it is a long way for them to. So we try to keep it on a higher level to create ideas and then the CEO hopefully say we have an idea and they say to somebody in their own organization try this out, see if it works and then we get momentum in it. But <Project Information> is fairly new, it was it's just like <Number> year ago since we started working with the concept because there was an idea for the politics and so we, it was launched in <Date>. So we, but the network is coming there. We had a really good meeting in like 3 weeks ago and with over <Number> participants. So it was nice. So I'm thinking to take this. We didn't get that yet. Some of the companies in this network are interested in last mile logistics and it's stakeholders like transporters of course, expeditors, it's the shops and the businesses in the city and it's actually it's also energy companies because they see we need to do this in an electrical way. How do we charge the vehicles and where do we charge Them? and what was I thinking and it's like it's companies like <Company> that are collecting garbage normally, but they can see if you just take the The dry garbage, you can combine it with other transportation, for instance, and because they're driving all over the place, and also the mail <Company>, can we find synergies? And it's really interesting so, so, so we are planning to have smaller more focused because when we meet like <Number> people it's it's very different differentiated companies that some are like <Company>, but they're very interested in the poor and questions like that. While as <Company> is very interested in what is happening in the city. So we're going to have smaller groups talking about issues and there We want to introduce the digital twin to see if we can get the actors, the different stakeholders to pitch in with data and engagement.	DS	
00:22:00 SPK_1 Yeah, so for like designing digital twins of this one, what do you think is the most important things in your mind about this?		
00:22:18 SPK_2 What is the most important things? Yeah, well,		
00:22:22 SPK_1 That's a big question.		

00:22:23 SPK_2 Yeah. It has a lot of different angles to answer that,		
00:22:31 SPK_1 Yeah, That's why we are interested in your angles		
00:22:36 SPK_2 For me to make this work, it's the data, obtaining data, but of course. It's just as important that somebody reliable can handle this data and put it into the digital twin in a reliable way, in a way that mirrors the reality. So technical competence and. And I should say that this digital twin is a part of a bigger project, yeah, that we just kind of pitched into <Project Information>. You can Google that and you'll find some information about it. <Project Information>. That's why we could do this with a fairly small amount of money, because they already had the skeleton and then everything was there. So we just had to put in <Location Information> in it. And it's a different focus on in different cities <Project Information>. For instance, and make it into, But we are interested in packages, or public transportation For packages, yeah.	DC DSU	ED
00:24:15 SPK_1 Yeah. <Project Information>. Do you think that's because they're focusing or that's why they're focusing on that or why		
00:24:25 SPK_2 I couldn't say, They have different challenges. You can get a name from, I think <Education Institution>, maybe. I think it was. Yeah. And you can talk to. Yeah. You can get that name from here. And <Name>, you know <Them>? <Location Information>. Yeah. Yeah, <They> knows about it. Yeah. Okay. Yeah.		
00:25:11 SPK_1 Cuz I was thinking, you talk about representation so that the digital twin is representative of the reality is, I don't know if it's too early in the process, but is there some kind of consideration? Like how is that considered in the process of designing to collect data that is representative?		
00:25:38 SPK_2 Actually it's just like as much as we can Okay, it's sort of like that. Yeah, really we try to collect all the data we can get our hands on because they, There are so many different actors in the city, I mean you have the big ones like <Company>, so but then you have <Company>, and smaller ones <Company>, No, they're not that small, but they do have direct deliveries, from day-to-day. If an architect need to send their to a building site, they call. So it's like taxi for.	DC	PD
00:26:43 SPK_1 Yeah, Okay. So when collecting data, So what do you think is the difficulties about it other than there is various actors involved?		
00:26:55 SPK_2 The big difficulty with collecting data is fear. I would say in the companies that they will reveal sensitive data for their competitors or. Or customer data that is sensitive. We say that we don't need, we don't need the final address as long as it's zip code that we can randomize within that zip code because then it's fine because it will be randomly provided within that zip code from day-to-day anyway, so. But there's	DC DS	ED MD

some way to go and you know the big ones, <Company>, I think is fine, but the international ones like <Company> and <Company> they are very, they are more From the top down it's strong hierarchy. Yeah so they it's fear. So we need to eliminate that fear with knowledge and to give comforting in that it's there's no problem.		
00:28:27 SPK_1 Do you think that there's a mistrust in the technology in itself or. Where do you think the fear is?		
00:28:35 SPK_2 No, I don't think it's a mistrust in the technology. I don't think that's the problem because they, it's just a fear that the data will end up somewhere where it should be and they could be misused. And it's also an organizational problem because. Like I said, it's the people I talk to here, They're fine, they say sure. But then they ask a boss, that has to ask a boss and then it's too far away. So that's, I think that's the main difficulty, yeah, in obtaining the data. And well <Company> don't want <Company> to know how they are optimizing their routes. So and that's I think that's the main reason	DSU DS	ED PD
00:29:45 SPK_1 Okay. So they don't want to have a chance or risk that sharing their business secret to others because of the data?		
00:29:55 SPK_2 Yeah, and The transporters are stuck in models and structures that are created with not with CO2 in mind. But the system are so large and so complex that they are very hard to change <Project Information> No, it could be a reason.		
00:30:44 SPK_1 So why do you think that could be because of money or why could that be that? They wouldn't wanna.		
00:30:51 SPK_2 It's all about the money. I would say and it could be someone who's lazy and doesn't want to work, but I think as, we just have to be patient because like I said before, the CO2 currency is getting stronger so it will soon be about money to actually do something. So hopefully		
00:31:26 SPK_1 Yeah. But how do you think that actually policy is affecting the development of digital twins? Or do you think that it's actually favorable to develop digital twins? As you said, CO2 currency is increasing.		
00:31:46 SPK_2 I see the use of the digital twin as a way to do experiments. In a lot cheaper way than in reality, because you can do, I mean. Like I said before, the logistic systems and the models and structures that are there, they are extremely complex because they are all connected. So we can't just change in <Location Information> because it affects everything else all over. So if you want to do something, for real, it's really expensive and it takes a lot of work with the systems and the structures To do it so,		

So you can't just, well, let's try it. That won't happen. So you have to have a really good Beslutsunderlag,		
00:32:48 SPK_1 decision making, Solid ground. Solid ground.		
00:32:52 SPK_2 Yeah. I have to be on solid ground. And to have and you have to have many different things that shows that this will actually work in real life. It might not anyway, but at least you have to do all the things you can to show that it should work. You can't just it's, well, it's too hard to do the experiments in real life. So you can, It's kind of trying to make a shortcut to better system, that's the main purpose I would say to do a digital twin		
00:33:47 SPK_1 visualization and communication,		
00:33:50 SPK_2 yeah, yeah. And the ability to make different simulations to insert a solution and see what happens. If we do like this, the flows with different like this and we would get the traffic would go this way instead and that would be good for the city to know for the planning. Is that a good way to go or is it a bad way to go so you can examine different solutions In a theoretical way of course. But still, it's better than just trying it out. Then you have to do parallel with this. You have to do real life tests of course. Anyway, in a smaller scale like autonomous vehicles and things like that, you need to try them in real life to see how do other traffickants react to them and what happens. You have to do that when you put new elements into the environment. You have to try that out. And I think this digital twin is actually open source coded. So, right now it's a company in <Company Information> that has developed it because they are connected to this <Project Information> project. But any time we could actually take the whole code and just continue on our own if we wanted to put developers into doing that for the digital twin, yeah,		
00:35:40 SPK_1 And then you add your own data.		
00:35:43 SPK_2 Yeah, we could do that, right now We don't. No <Company> it does it for us.	DS	
00:35:50 SPK_1 OK. But are they, the data, is it from <Location Information> or is it.		
00:35:55 SPK_2 It's from <Location Information>, The data we deliver to them and they put it into the model.	DS	
00:36:02 SPK_1 So how do you collect data? You reach out to companies and..		
00:36:07 SPK_2 Yeah, so far we just took the data. I mean <Company> for instance, they know Very	DQ	ED

much about how much packages are delivered and then, So we put that statistic data into, at this point and we tried to verify it by, You saw all the different blue dots on the map, They are actually traffic measurements. So we tried to match it to see if it's, Is it correct or not. But, So the statistic data just for packages, e-commerce packages this is. But then we need to put in all the towels and sheets for the hotels and all the shield transportation. It's a big it's a long way to go before we could actually use this for last mile transportation overall for the city, but we have to start somewhere. And right now it's like I said, it's statistic data, with some exceptions because we got some <Company> data, but we don't have the other ones and we don't know if a package is traveling from <Company> In <Location Information>, we have. We have the start and we have the origin and we have the goal address, but we don't know which points it will take. So it just will go. And that doesn't happen in reality. So it's, it's a long way to go.		
00:38:00 SPK_1 So you need more data, basically.		
00:38:01 SPK_2 We need more data. Yes. And some more money to put the data in, but that I think if we got the data it's not a problem to.	DC	PD
00:38:13 SPK_1 So you mentioned about the <Company> in >Location Information> that developing, helping you guys develop the digital twins. So what's the decision making at first that choosing them and then to what extent you trust about the company to develop the Digital Twins for you?		
00:38:29 SPK_2 Actually we just docked in to the project <Project Information>, And they were there, Okay, simple as that. We started this out by meeting with all the stakeholders like two years ago and we had a consultant that interviewed them and it ended up with a suggestion from the consultants that we should gather the data. from all of them and put the data with layers on top of each other and try to see if we can make a different transportation plan. And they gave a business proposition for about <Number> Swedish kroner to do this, its like Oops. And then <Name> from <Education Institution> said Have you seen the <Project Information> project? Maybe you could dock into that. If you asked me. And we did this for a 10th of that amount, we could get this started. So. And it's actually. The same thing that we want the digital twin to do, to take all the data, put it on different layers on top of each other to see Okay, if we do it like this, it's a smarter way. And that's the main goal for. But we have to take one data point at a time. So the <Company> was there and they knew the product because they had developed it. So we just asked them, put <Location Information> into this, how much does it take? And I applied for some money from our innovation fund and I got the money to do this. So that's basically what happened.		
00:40:27 SPK_1 I was thinking, do you think there are any concerns when, cuz the companies are collecting data for one purpose from the beginning? And then you use it maybe for another purpose or goal. Are there any considerations in that transfer? Transformation of the data		

00:40:48 SPK_2 It could be, of course, because when one business take, export their data they have, it will show up in a certain format, yeah, and when this actor. Export their data, it comes in a different form. And when you put them together, you of course have to be very careful not to sort the apples with the pears. So that's a process, of course, and it could be, of course, but it's actually. That's the main thing, not to sort apples with pears, but there could be issues where we don't know the difference. of course it could happen, because it's like when you hold your nose and you bite an apple or. You can tell the difference, Or even a potato. We can't tell the difference But we have to be very careful. It's not. I figure it's not really that complicated. Goods traveling from this point to this point, to this point, to this point and ending up at a certain zip code. It's not really that complex of data. The data is not that. So we should be able to sort it out I think.	DQ DS DSU DC BD	ED
00:42:53 SPK_1 Did you bump into any complications with this so far?		
00:42:07 SPK_2 No, I think it's very, they're very smart.		
00:43:16 SPK_1 You're trusting the IT, Yeah. And the technology. How would you say that your knowledge of the technology is do you feel confident?		
00:43:25 SPK_2 My knowledge of the technology in itself is very low, they are. Because they are actually writing code. And I'm not into that. So I have to ask them everything.		
00:43:44 SPK_1 And you think they like you have a good communication?		
00:43:47 SPK_2 Yes, very good. they are very lyhörd. That's a hard word. They're very, I got the word, They hear what I say. And they adapt. And they try to understand. <Project Information> The mindset.		
00:46:02 SPK_1 So talking about this is kind of relate to your question that we got, is that about a project success and failure? Is there anyone like responsible for the success and the negative outcome of the digital twin projects? Who do you think they are? Because there are so many actors in the project involving.		
00:46:30 SPK_2 I think you have to understand from the beginning that it's not an easy thing to do to make it realistic because everything you put in, For instance, like last mile logistics, every four years about when the politics change in the <Location Information> there is someone new coming up with it, "well, we should consolidate all the goods outside the city and then so everybody delivers it there and then it goes down to the city, It would be a lot smarter way" and it sounds really easy, yeah. But when you scratch the surface, it gets more and more, more complex. And we have to show them that every four years about, that is more complex than you would think of the first glance. And that's the reality with I think any digital twin as well, because reality	BP	

is so complex. Because it has to do with people. And people are complex. We are not logic creatures. We do things from different. We're not logical always. we do things from emotional point of views and to put emotions into the digital twin is a lot harder. So depending on what kind of digital twin, because like this twin for instance, it has to deal with preferences in how you want to receive your packages. we don't really know what's gonna happen. And what possibilities that will show up I want maybe I wanted want someone to deliver me something to drink while I'm out running. I don't know. It's, so I think you could conclude it with that the difficulty to include people's behavior into the Digital Twins is not to be underestimated.		
00:49:08 SPK_1 And these preferences for delivery, for example, are they somehow included in the data right now?		
00:49:14 SPK_2 No, no.		
00:49:15 SPK_1 It's just the start and end point of the delivery so far.	DC	
00:49:18 SPK_2 Yeah.		
00:49:19 SPK_1 Is there a plan to include more of the complex in this model?		
00:49:26 SPK_2 In this model they would show up as delivery points I would guess. So you kind of, you kind of you take the preferences and you make them static.	BP	
00:49:41 SPK_1 A time and a place,		
00:49:44 SPK_2 Yeah. Yeah. Because otherwise of course because <Project Information> And if you really were AI, the system would be able to learn behaviors and things like that. I don't think they're quite there yet.		
00:50:13 SPK_1 It's nice to start the AI topic because we want to know, do you have heard about algorithmic bias? These terms, wat is your understanding about it.		
00:50:26 SPK_2 I would guess it's if I'm just making guess, I would guess that AI is actually created by someone writing code and their points of views are kind of adopted into the code. If that's what you mean.	BP	ED
00:50:52 SPK_1 That's one part of it. Yes, that's definitely correct. And then there's also the data, the data in itself can be biased as well. So there might be more of yes right or in one area and then it makes more accurate predictions about that area. And		

	BD	ED
00:51:09 SPK_2 so if the data includes that people hired on a certain position are at 95% Swedish and called <Name>. The, the, the computer would think that the AI would think that that's the way it's supposed to be.		
00:51:31 SPK_1 Yeah. Yeah. So that's who we should hire for these positions. Yeah, exactly. So that's I guess we were wondering if that's like something that's considered somehow when collecting data and. And how you when you're looking because you said you're also looking into the data and also do some analysis on it. So once if there is this kind of bias there, how do you think that you will try to like rectify the situation or do something else? I don't know. We're interested to know.		
00:52:08 SPK_2 It's a good question because I don't transportation data is not. I could be wrong.	BP	
00:52:22 SPK_1 Nobody's wrong here.		
00:52:24 SPK_2 But at the first glance you would say it's not the biased data because it's all about where it's supposed to be delivered. But then of course. If you would connect the AI into it, the AI would get the information that very many people like to have their deliveries to <Company> over there. But that might not be the truth because <Company> is an ombud where you collect your packages. But actually they would if it was possible, They would want it to be somewhere else, things like that could happen.	BD	
00:53:17 SPK_1 That's a good example I think because it could also be maybe that the data is more centered somewhere and then those people like to have it in <Company> or like they they pick up their packages in <Company> and then the system might optimize towards that and then people outside of that area might somehow be negatively affected some how.		
00:53:41 SPK_2 Yes, But I think the whole purpose of the digital twin is to simulate differences from what it's like now. So what happens if we say no more ombud? What would happen then? And you want to question that.		
00:54:07 SPK_1 So you're trying to deviate from the standard way of,		
00:54:11 SPK_2 yeah, trying to give it a blank paper and see what's the best way to do it. That's what we would like to do.		
00:54:21 SPK_1 So then coming back to the citizens in this specific example, would you see a value in them, including them in the discussions?		

00:54:34 SPK_2 Absolutely.		
00:54:35 SPK_1 Yeah. Is that something that's in the plan or		
00:54:40 SPK_2 no, we don't have that in plan for this digital twin at this point at least, but then in the other hand we everybody who's connected into it are also private citizens. that order things on the Internet and we can kind of sort of relate to being just a person, I do that all the time. I mean, I don't think Nah, I would never wanted that as a private person. And then of course you have to ask, but I think many enough to sort of maybe that's good for you but maybe not for you depending on life situations and where you live and where you work. How you move around?		
00:55:47 SPK_1 Yeah. Cuz that's also interesting, cuz one kind of bias could be having then more opinions from people with a certain lifestyle.		
00:56:00 SPK_2 Yes. So that's the risk of course that we are quite similar, the people at this point that we behave similar. And that's a risk everywhere, not only in Digital twins.	BP	TCA
00:56:21 SPK_1 Is this somehow a discussion in the project or is it		
00:56:30 SPK_2 not from our point of view at this time because we're just not there yet. Maybe you should ask that to the <Person> on the <Education Institution> how they're thinking when they when they are trying to simulate <Project Information> because that's a lot closer to. To to at this point. Yeah.	BD	TCA
00:57:04 SPK_1 Yeah we can we can get the name maybe later.		
00:57:11 SPK_2 <Project Information>		
00:57:45 SPK_1 Is it one of the company that we have contacted before? I'm not sure. I don't recognize <Name>.		
00:57:55 SPK_2 <Location Information>		
00:58:00 SPK_1 OK, we'll have a look at that. Yeah. And. So talking about the data that collected, the statistic data collected and so who is the owner of the data. So after you guys collected them, is it considered possession of the municipality or actually still consider is a ownership is from the original private company or something?		

00:58:28 SPK_2 I think we can't say it's our data. I don't think so, And I think to get the data from the stakeholders we will have to sign agreement of secrecy, and to use it, but ownership I don't know.	DSU DS	PD ED
00:58:55 SPK_1 So the companies that own the data, just lend it to you.		
00:59:03 SPK_2 Yes, I would think so.		
00:59:09 SPK_1 I'm just wondering is the kind of question that about future that if the Digital Twins is going to like implementing in the future like all the data collected used in the digital Twin, will there be any consideration to open to the public?		
00:59:27 SPK_2 I Why not? If as long as it's anonymized the data in itself, it's anonymized and you can. I don't see at this point, at least not for this thing, what the public would gain from and be interested in it, but. There could be other things that would be more interesting. I know they're trying to build, they want to build a digital twin in 3D for a whole part. 3D city. And that's extremely complex, I would say, to get the people to behave and that would of course be a that would be fun to experiment what would happen and if you could just, fantasizing about what would be, what would happen if I open a cafe here, what would be the flows around? if you could simulate that would be very interesting. So I don't know if there would be interest for the public maybe.	DS DSU	MD PD
01:01:08 SPK_1 I would be interested.		
01:01:10 SPK_2 Yeah. Well, you as a public or as a student.		
01:01:19 SPK_1 I'm a student and also a public. I was thinking, are there any concerns? Because I'm thinking if the data would be identifiable to a person, then that would go under GDPR, I guess? Or how does that work?		
01:01:35 SPK_2 yeah, If it's possible to trace the data for a specific person,		
01:01:42 SPK_1 yeah. So if you can identify a person		
01:01:45 SPK_2 I don't think we would want that, no. It's the same if we take pictures of if we use cameras to identify vehicles, I don't think we should register the ownership of the vehicles, just the vehicle type and pollution.	DC	PD

01:02:05 SPK_1 But are there any, except for the GDPR aspect of it, with personal identity, are there any other concerns to that?		
01:02:16 SPK_2 Yeah, absolutely. Because like in this digital twin, we say that zip code would be anonymous enough, but some has their own zip code, so you know, everything delivered There is just one customer, and that could be a problem. It's not GDPR, but it still could be sensitive data, and large companies the same if they have a plant or they have their own Zip code, And so the data would be identifiable.	DSU DC	ED MD
01:03:10 SPK_1 How do you manage that right now if that's already?		
01:03:13 SPK_2 We don't. No, because packages are. No, we don't. At this time.		
01:03:21 SPK_1 It's not considered like sensitive data?		
01:03:24 SPK_2 Not the data, not the data we have in there right now because it's based on public statistic data. And then we simulate it. So I guess you just have to, if it's a problem on a certain zip code, you have to make that square large. And to simulate between different zip codes within the area. And then, so then you can't know if it's really, if it's actually delivered to that point or to that point, yeah.	DQ	ED MD
01:04:03 SPK_1 Are there any standard processes or ways of deciding what kind of data is sensitive or what needs to be protected more is that?		
01:04:17 SPK_2 I don't think it's there's a standard except for GDPR,		
01:04:21 SPK_1 Yeah. So it's more a discussion then		
01:04:25 SPK_2 yeah, I would say so what's yeah,		
01:04:30 SPK_1 between the stakeholders?		
01:04:33 SPK_2 Yeah, and depending on what the goods are. I mean if it's medical, it could be sensitive if it's just. Take. I mean, just so who the customer is and what the content is, it could be more or less sensitive.	BP BD DS DSU	ED MD

01:05:00 SPK_1 and that's a discussion. Then from time to time?		
01:05:01 SPK_2 I would say, yeah, it would be that. Yeah. Yeah.	DSU	TCA
01:05:05 SPK_1 I was curious about you said that if there's a problem with a specific zip code, if that's too. Identifiable and you make it bigger. Is there then risk that the data is too simplified to?		
01:05:22 SPK_2 yeah, I could be, theoretically yeah. Because if we go back to this twin where it's actually last mile transportation, if you have a zip code here and the only entrance is on this side and you make it a bigger square and the streets over here are one way in that direction, you might have a completely different routing to reach different. So that's just a simple example, but it could. Effects like that could of course occur. But then it's a compromise that you have, I think you would have to live with.	DQ	PD ED MD
01:06:20 SPK_1 But how? Sorry, I was just thinking that that's also a discussion for that specific problem.		
01:06:27 SPK_2 Yeah, I would say so. Because and you could identify this problem, this will show. There will be a problem with this data and this simulation. So how do we get around it? How do we do that? I suppose you have to handle those manually, problems like that.	BD BP DSU DQ	TCA
01:06:47 SPK_1 And then with decision making it's then a compromise to reach consensus. Or are there any hierarchy for for who has the last say in things or in the discussion?		
01:07:00 SPK_2 the data owner of the data owner I would say. We can't do anything with the data without there.	DSU	
01:07:18 SPK_1 I feel like we got answers. Yeah, I think so. Yeah. Do you have. I don't think I've got much. Sorry. It's past like 7 minutes something, already. No it's just one minute. Yeah. We started early. Yeah, I guess. Are there any other limitations or concerns that we didn't discuss yet that you would like to put further?		
01:07:55 SPK_2 Not that I can't think of right now.		
01:08:04 SPK_1 Or anything in general, maybe not a concern is there's something		

01:08:07 SPK_2 no in general. Just like we talked about the hassard of underestimating complexity.		
01:08:22 SPK_1 and is that something that you actively work towards?		
01:08:28 SPK_2 No. Well, it's something that we keep in mind, that looking for looking out for too simple solutions and answers because if we find an answer really easy. Naah. I think one of the strengths of getting a network with all the different kinds of stakeholders is that you also get all the different kinds of perspectives, and I think that's very important because if we as the <Location Information> starts to talk with the transporters only, For instance, we can find perfect solutions for the two of us, but it might not be perfect at all for some of the others. So and it's kind of cool that the companies and the businesses are, They say yes, we understand we need to cooperate. To find the solutions. They accept that. And that's a very good start. It's a very good start.		
01:10:04 SPK_1 And so you're trying to include as many as possible to reduce the unknown complexity, kind of,		
01:10:09 SPK_2 yeah.		
01:10:11 SPK_1 That's interesting. Actually I got a question. When you guys talking about the grid and the zip code and the routes. So when doing a digital twins, how do you think that all the actors, stakeholders, they will weight the performance of digital twins like the accuracy or they will value something else like the zip code like and larger zip code, so it's more anonymous and then not to be identifiable. So which one do you think that they will weight more when talking about this kind of Things? because I think it will be interesting to.		
01:10:49 SPK_2 I think it will be different from active tracking. I think some of them will just say here's the data, That would be the smaller ones that don't have to ask someone in <Location Information>, and some of them, It will be a long process and they will and it also depends on who the customers are and. So I think it's could be.. I think maybe it's like this that sometimes the transporter will have to ask their customer if they could give out the data	DS	ED
01:11:39 SPK_1 to get consent. Yeah, yeah, Okay. Is that a problem in general with consent from the people or the units that are included in the data?		
01:11:53 SPK_2 Not right now, because we don't have so many actors in it, but I think it will be because it could be, for instance, a producer that if people knew how many deliveries we'd get, they could go calculate that. How big is the production at this time? And, you know, concerns like that, it could be showing up like Gubben I lådan. We don't know yet. Yeah, it's difficult. There's so many things you want to do, But	DC DQ	

again, the complexity, you never know what you will bump into that is a factor in collecting the data.		
01:12:56 SPK_1 It's interesting to know your opinion. Yeah. Well, is there anything you would like to ask us or?		
01:13:09 SPK_2 No, I'll be interested to to read what you come come up with.		
01:13:21 SPK_1 Yeah. And I was thinking if there's anything that we realize afterwards that we forgot to ask or it's okay to e-mail or call.		
01:13:32 SPK_2 Absolutely. Yeah.		
01:13:34. SPK_1 Thank you. Thank you. Then I. Then there will be. Yeah. Yeah. So thank you so much. Yeah.		
01:13:40 SPK_2 I will e-mail you right away because then I'll forget it. <Name>.		
01:13:49 SPK_1 Yeah. It was a very interesting discussion I think, yeah. And then different people have different project, digital twin project for different purpose. And it's so interesting to know we should maybe look into, I guess we can.		

Appendix 6: Interview 4, P4

Theme	Colour Codes	Factor	Text Codes
Context	Yellow	/	/
Data Management	Blue	Data Collection	DC
		Data Storage & Use	DSU
		Data Sharing	DS
		Data Quality	DQ
Algorithmic Bias	Red	Data	BD
		People	BP
Principle of Discourse - Democracy Principle	/	Pragmatic Domain	PD
		Ethical Domain	ED
		Moral Domain	MD
		The Theory of Communicative Action	TCA

Participants	
Authors	SPK_1
P4	SPK_2

Transcript	2nd layer	3rd layer
00:00:02 SPK_1		
I guess, yeah. And it's working. I think you have to recording in progress. Yeah. Thank you. So do you want to start by telling us a bit about your role right now?		
00:00:26 SPK_2		
Yes, I'm <Name>. I work at the <Company> at the <Location Information> as a <Role> for many different kind of innovation projects. And everything is about visualization. That is our main focus, to use visualization for different challenges. In society in many ways, and my focus is digital twins for cities and regions, so that's my main focus.		
00:01:06 SPK_1		
Do you want to tell us a bit more about the digital twin or twins that you're currently working with?		

00:01:12 SPK_2		
The digital twin that I've most been working with is the one from <Location Information> <Project Information>. And I don't know if I sent you the link, but as the beginner that <Location Information>, they have <Number> videos about the digital twin. And so if I haven't sent you the link, just remind me that I should because they had they described very, very well what they want to do with the twin in those videos. And I'm not actually involved in the work with the digital twin. I'm involved in a <Platform> that is connected to the digital twin. ==So our focus at the <Platform> is, is trying to develop the usability of the digital twin.== Because when you are creating this new technology, it's sometimes ==it's hard to understand how you should use it and even if you know that. What you think they will be interested in using the digital twin== for it's a very long ride. Just try to make your colleagues and politicians and other people ==to understand the potential in working with new technologies. So there is more of a bit of a work for change== and so so that's the focus that I'm . My focus is to create change by using the digital twin.		PD
00:02:48 SPK_1		
Okay and what are the organizations and companies people that are participating in this project?		
00:02:58 SPK_2		
In the <Project Information>, we created <Platform>. I don't know if you saw it. Have you seen it in our website? <Project Information>. That's the <Platform> that I've been working on. I can, I can send you that link as well because during two years we were like ==<Number> different partners== working to try to understand how should this <Platform> work. And there were partners from the city, of course, but also from <Education Institution> and <Company> and <Public Institution>? They work with education and different companies like <Company> and the and also small companies for one two people working very in in very small companies and they were working with the gamification and they were working with different. So there was a broad variety of people from small companies to big companies and the city that were trying to understand how could we use the digital twin.		
00:04:25 SPK_1		
Okay, and I think you mentioned it a bit before, but what would you say are the overarching goals of this project?		
00:04:36 SPK_2		
I think the overarching goal was to. How can we use? Actually, we had two questions that we wanted to answer. One was how can we use the digital twin and how? How can we work with urban development by using the digital twin. And the other question were how can we? How can we? What do you call it? Tillgängligare?		PD
00:05:12 SPK_1		
Make it accessible.		

00:05:14 SPK_2		
What do you say?		
00:05:16 SPK_1		
Make it accessible maybe?		
00:05:19 SPK_2		
Yes, maybe, maybe. I think we had a nice word for that in my PowerPoints, but it's to have the let's see. I think I can have the nice word for you here. I can share my screen if you want to see how can we make the digital twin available is one question. And the other question is how can we work with urban development by using the digital twin. So I can just share my screen if I can, yes, I can share my screen and you can just if you want, you can see that. So these are the two main objectives that we were working with. so. In the <Project Information>, this is in the <Platform>. But the reason why we created the <Platform> is that <Location Information> when they were developing the digital twin, they realized that if we just develop a twin. ==Then nobody knows how to use and nobody knows what to put in it. We will surely fail in in have. This Usability is very important if you want people to finance your project. It's very important that you have a good result. And if you want a good result, you need to understand this new technology.== So if you want, you can just make a screen dump or if you want me to send it to you, I can send this picture if you want to.		PD
00:07:23 SPK_1		
Yeah, I think we we got the questions in the recording. So may I ask because the project is all about the usability, so is there any use case that you can describe some of them to us so we understand more about it?		
00:07:39 SPK_2		
Yes, we had. We started the <Platform> from 5 societal challenges, 5 challenges that the city was struggling with and we tried to use the digital twin. We had short sprints for three months. We just we took a we took a challenge and we said how can we use work with this challenge by using the digital twin. So the first one was flooding and Skyfall heavy rains. What if we could use the digital twin to to be sure that if we are, if there is a flooding or a torrential rain, we will manage to keep functions like the train traffic from the they are building a tunnel. Under <Location Information> for train traffic and they want to be sure that if there is a lot of rain, the train is still able to go. So that was one question that we were working with and we made a small demo or pilot around the question and then we had questions about as social questions like a citizen's connection, how can citizen be able to to work with the digital twin. To cocreate an area and then we've made an augmented reality application where you can work together with your phone. ==You can go out in the city and you can work with your phone to to create something together. Have lots of people making comments and so on on on a place.== And so that was more the the societal side and then we had self driving vehicles, because the self driving vehicles will change the way you plan your city and the way the city works. So we try to to make a pilot of that question as well. When the when you put autonomous	DC DSU	TCA

vehicles into something that they call shared spaces, it's where you have people walking, you have. Bikes, you have all the different kind of of buses and everything. So we try to understand how could the digital twin be a place where you can test these autonomous vehicles. So there were three and then we have outside <Location Information>. They are going to rebuild the town a lot around the because they are going to to build a lot of new buildings. On on <Location Information>, so suddenly there will be a lot of traffic coming from the other side so so they need to to reform the place in front of the central station. And that was traffical questions we were looking at because <Public Institution> in <Location Information> they wanted to be able to dynamically work with the place. Because when you are looking at people walking, you do a simulation and then you have the result. And if you want to change something you have to do the simulation all over again and then look at the result. They wanted to make changes and the people in the 3D model should start to react on the change and so we made something called <Project Information> or the people in the model, they understood when we moved the zebra crossings or when we moved fences and so on. So you can move fences and zebra crossings and the people with changing the way they are and they also changed their behavior according to rain, because that is something that you don't get from making simulations. You don't have the weather as a as a factor and in <Location Information> when the weather, when the rain is is really a factor that people are, you know, when it's raining people start looking like this and they crossing the street in different ways and they running and so on and so they don't behave as they should. So it was interesting to see, can can we create this other kinds of behavior to see when it's sunny people walking like this and when it's rainy. Maybe we should have that in in in our minds as well when we are making the plans. So that was the 4th about the traffic and this, the last one was called the the <Project Information> because <Location Information> is has been quite low and not so many high buildings and now we are building a lot of, lot of, lot of very high buildings. So that would change the whole way that the city looks. When you look when you're approaching the city from the sea or when you're coming driving from <Location> or <Location> or whatever, the the city will look very different. So and because of you're making a lot of changes in a short period that you you want to know how will the city look like in 50 years. So so even if it's not everything is not already built. Because if you have a city and want to make a small change it's easy to understand how the the city will look like. You can have you can make this photo, photo montage and see that okay this building is going to be like this but when you have a very huge areas that is like just flat. Then you don't have anything to relate to. So you need to have, you need to have the the time slide when you can see that okay this is the one, this looks like this today everything is flat and then we're going to start building and how will it look like in 50 years. So, so that is was the the 5th pilot so and that gave us a lot of information about. What the digital twin could do and what it couldn't to do, So so I think that's a good way of like crash test your digital twin. Can it do the things that we need it to do? Because if it can't, maybe it's the wrong tool or maybe we'll have to change the tool to. Because we we need to work with the questions that we are working with. If you're working with flooding, you need to have a tool that can help you with flooding, not anything else. So and it's very important that all the people in the city comes together and see the digital twin and the effects that it has. So I think the digital twin is a very good tool for this, but. I think it's a long way for municipalities to change the way they are working.

00:15:23 SPK_1

Hmm okay cause I was thinking about the different goals than different stakeholders might have slightly different goals. So how do you reach a consensus regarding goals? Like what's the process for that?		
00:15:41 SPK_2		
I I don't think you can ever reach any consensus because the the goals are always and they are always when you're working with the city. It's it's a small area with a lot of lot of lot of interests. So there will always be a conflict of interests and there will always be. A lot of different goals, so, so maybe the common goal it could be to make the city as good as we possibly can and try to make the right decisions. Because everything with the working with the city, it's it's always a conflict of interest and you cannot you you can never. Satisfy everyone so but maybe you can use the digital twin to to to see the whole picture. So you can even though you cannot have what you want. Maybe you can accept that others need this and this need this. For instance if you if you want to have good transportation with cars that there will be noise and if you want to have. Good. If you want to be able to build high buildings, it will be darker streets and it's always like give and take, give and take, give and take. So maybe the goal with the digital twin is to to have the common understanding to to use the twin to to build as good city as we possibly can. But, everyone doesn't have the same goal.		TCA ED MD
00:17:40 SPK_2		
So that will be a conflict, yeah.		
00:17:43 SPK_1		
Is there any hierarchy or how do you weight the goals? Can there be goals that are weighted higher than others?		
00:17:55 SPK_2		
I think when you it depends. I think that depends on if you're talking about what the digital twin is. Or if you are talking about what the digital twin does, because when they are creating the digital twin, they have a like a 10 mile long list of interesting things to do and interesting things to put in the twin. And they are limited of course in their possibility to put everything in there. So they need to. To make priorities. So for instance, maybe you prioritize to have all the trees in the model. All you prioritize to have real time traffic in the model and then you just list everything and then you start from priority one and put it in as long as you have the the resources and the money So and and who is doing that prioritization? When you can discuss how? Who is? Who is doing the list and who is doing the the prioritation? That's the question that the city is working with, Yes. And of course it's about resources and it's about money. And so they have their list and then you can always question if that's the right or the wrong, but you don't know. But then you have the prioritation. Of the people working with the digital twin and now I'm talking more about the people making decisions for the city. That's the politicians. How, how the the politicians prioritize both the work with the digital twin but also the the things that come out from the digital twin, are they using it to make decisions for the the city or not So there are lots of priorities always working around. So it's like but if you're working with if you try to work with something that is very urgent matters and I think the		

<mark>climate questions and all the different kind of questions around the environment and sustainability is, of course, very important.</mark>		
00:20:42 SPK_1		
So going to going back to the one of the use case that you mentioned about the social challenges regarding to the citizen and also their behavior and something. So is there any like participant that they can actually participate in the digital twin project and provide feedback to you guys, are there any citizen involved in the project as well.		
00:21:06 SPK_2		
Not in the <Platform> yet.		
00:21:07 SPK_1		
Okay.		
00:21:10 SPK_2		
And and I don't think in the digital twin but <Location Information> has had for many, many years something that they call <Project Information> and it's an application where all the citizen can go in and just make comments and and they can create things on their own and they can. So, so I think that is like the the 1st way they're starting to work with digital innovation but that's not really on my table. So I will not, yeah I will not say too much I tell you something wrong but but the city <Location Information> is very interested in using these technologies to work with. All the people in the town both the ones who live here but also the one who comes here for instance the all everything has to do with tourists is very interesting to show in the digital twin as well and and the people working here because they want to to be able to open up the digital twin for for business, of course there are a lot of companies wanted to work with the digital twin, like architects or the car companies or infrastructure companies wants to work with this kind of information. So they really want to open up.		
00:22:48 SPK_1		
Okay. Thanks. Did you have a? Yeah, I got a really broad questions that I want to ask. So what do you think is important when designing or implementing a system or digital twins that used the for city planning?		
00:23:04 SPK_2		
Use the.		
00:23:05 SPK_1		
For city planning.		
00:23:07 SPK_2		

For city planning, yeah, I I think it's the most important is to start with a challenge. That you really need. I think that is the most important thing. Maybe the second important thing is to don't wait for everything to be perfect because it will never be. So just start with your question or your challenge and try to solve that one and see okay was this good or bad. Whether we take this further or not, how can we change? So you work like in small iteration, so otherwise you will just drown because it's getting so big. So I think it's important to to have it close to you and to work in a small step so you can so you don't get like overwhelmed or you have to have everything in order because you will never have everything in order so. I think that's the most important thing maybe.		
00:24:17 SPK_1		
OK. Thanks.		
00:24:18 SPK_2		
I was thinking about something else as well. Yeah, maybe I maybe it pops up.		
00:24:28 SPK_1		
Yeah, we can come back to it, Yeah. And how would you say that policy making and policies in general are affecting the development and implementation? Of the digital twin towards the goals.		
00:24:44 SPK_2		
I think policy making is very, very important in this because it's always easier if you have somebody else to hang on to and and try to you, you know, you say that you shouldn't. curla dina barn. You shouldn't make everything easy for your children because the children needs to struggle as a parent. But I think sometimes working with the city development and the urban development and digital twins, you feel that OK. It would be nice if somebody just clean the road a little bit. So there are some questions. ==That would be interesting if you could take them on a higher level and take it on a national level to to be able to learn from each other and maybe put some regulations in in place as well.== I I don't want to go into exactly what, but ==there's been a lot of discussions about the national hub for Digital Twins and this national hub==. Could it could be a way of of gathering all questions and have a a way of talking to the politicians like with one word and maybe it could make it easier to to work with digitalization and visualization and digital twins on a higher level because I I think we. We have a lot of people working with digital twins and we really could have some support in different ways as well.		
00:26:46 SPK_1		
So this is then to promote this work I guess. Are there any limitations as well from the policy perspective?		
00:27:01 SPK_2		
Well the the, the policy perspective always needs to really keep the what do you say to set the set the standards high They they cannot just like go and change things	DSU	ED

from day-to-day and back and forth. They they need to be they need to be a little bit slow just to think And try to understand, but sometimes I feel that they could do maybe temporary, just as when you're working with digital twins, you're working with a lot of testing. We have a \<Platform\>. We try, we try out and maybe we could have like regulations tested in the \<Platform\> as well. Like, what if we could do this, What would that? So maybe you can start like trying testing the regulations in the digital as well, so you can see the other okay, if we could do this, what could that lead to and then it could be time, it could be like. Separated like a parallel lane and the very specific regulations around what you are testing and then you can be safe and feel safe as a politician to say that okay, this is not a new regulation but this is like a demo test area for regulations and then you can go banana see that in that area because then you can feel a little bit more free and I know some, some governments are more like, hey we try this and if we fail we just go back. So they do it in in an open scale at once. But if you don't, if you feel a bit afraid of doing everything in an open scale, maybe you can use the \<Platform\> or something like that to try in the small scale. I think that could be very interesting and maybe we could get a little further. Quicker, but there are very important things, especially <u>when you're working with social questions. We have the integrity of people and the information and GDPR and everything that you really need to take care of the information that you present in a digital twin, of course.</u>	DQ	
00:29:40 SPK_1		
Yeah, I was going back to one of the use cases you talked about earlier with the behavior. During rain or not, yes and connected to this, how do you collect that data or what is that based on that simulation?		
00:29:56 SPK_2		
Actually, that simulation is just a pilot, So we don't have any perfect information. So it was more like test if we could make them change the behavior. So there are nothing scientific so that you can lean to and say this is the correct, but it was more of an idea to understand that this behavior is like missing. So that was more of an idea to test that it wasn't scientific. A project and and those avatars you I cannot say that they are working probably correct because it was most a way of visualizing that this could be the way to work because the pilot was just three months so you don't you don't develop you don't develop anything in three months so this was. Just a called it a sketch or an idea prototype. So it's a very interesting question. So I really, yeah, I would really like to see somebody make a very big project around that would be interesting.	DQ DSU	
00:31:27 SPK_1		
So it's similar to a proof of concept project and something like that. So is there any use of like real data in our world? Or it's just some kind of mock up data that you're using for the \<Platform\>?		
00:31:41 SPK_2		
All the data from \<Location Information\> was real data. All the data we got from \<Location Information\> was there. All real data. So that was everything, was the real twin and the real information. And we used real time data from \<Platfrom\>, the		

bicycles and that was the real data as well. So, so the data we work with were real but our pilots were like sketches. So just try to understand how could like the question was how could we work use the digital twin to solve our challenges so.		
00:32:27 SPK_1		
Oh it's very interesting to hear that the real time data that's why because we got some interview before and then everyone is talking about OK you may ask the <Location Information> project maybe they're using real time data so we know that it's using.		
00:32:41 SPK_2		
Yeah yeah, I think it's a lot of fuss around the real time data. I think it's very important when you start to use real time data that you need to understand. And the usability, what are you going to use it for? Because if you start to collect loads and loads of data and you don't have any idea what to do with it, you might as well don't do it because I think it's important to you. You can gather a lot of over data and say that maybe we need this in the future, but I think I think it's always good to have. Like a vision what you want to do with your data there's some kinds of data is very natural that you want to the traffic data for instance or or weather data or sound noise pollution, that kind of data data I think that's but still if if you are gathering data for instance about noise. Then I think you should know why are we gathering this noise data? Is it because we want to reduce noise levels for citizens or in a park or wherever? Are you going to use it to reduce the noise level in in a block of flats or whatever? Or are you going to use the data to to locate the areas that are still silent in the city? Maybe you should have in your planning that, okay we can see that this area in our city is the most quiet area in the city. Maybe you should put the regulation on that one and say we are not going to be anything. That make this spot more noisy. That could be. And then you have a purpose on why you are collecting the noise data. You want to reduce noise or you want to save and and and regard the silent places in the city. But just to collect a lot of noise data, what are you going to do with it? Just put it in the can and have it there for ages or what? So I think you need to have the purpose always when you start with. Because I think one of the most important things with the digital twin, it is the function that you can have the connection between the the city and the digital twin. So I think the real time data is very important, but I think also it's important what you want to do with it that you have a plan. If I have this data, what should I do with it because I have this? Yeah, I think it's <Location Information> was a very old definition of digital twins that that they wanted to use the digital twin to be sure that they that they got the project that they they needed the the, the, the product that they wanted. So if you want a good city, can you use the digital twin for developing good city. And then when you have build the city, can you then check that you got the product that you wanted? And if you didn't get the product that you wanted, can you go in and change it afterwards? Like if you wanted to build a city, a part of the city with the different good values or or? How can you, how can you test that in the digital twin in advance? And how can we check that you got the right result afterwards? And then it's very important to see Okay, we want to build this city area, it should have lights like this, it should have this kind of lot of bees or green areas and you have a lot of criterias. And when you have the criterias, then you build the city and then we need to follow up on all the criterias. Did we get this much? Birds or less accidents with the pedestrians or whatever you had for criterias. So then you need to have the	DC DSU	PD MD

information back to the twin and see Okay, we did get everything we wanted or we didn't. And if we didn't, how can we change? So I think that's my opinion on the on the real time data that you really need to know. Why am I collecting this? What am I going to use it for? And if I have data, yeah, what What to do with it otherwise? It's like, yeah, I think you understand, yeah.		
00:37:58 SPK_1		
So I was thinking then the, the groups of people and the areas that you're collecting data about, how are they being represented in the digital twin like are they, is there work, is there work towards making like a good representation of The data.		
00:38:23 SPK_2		
Do you mean any equality terms? Or do you mean		
00:38:26 SPK_1		
yeah, might be or yeah, if it's accurate, I guess.		
00:38:37 SPK_2		
Because I think that is something that the digital twin is a little lack of. It's the validation. If that's what you mean to to validate that if you are doing simulations in a digital twin and you get the result when you do it in the real world. So I'm saying do I have the same result or how much does it differ? That's that's one question I think. That is very hard. Maybe it's easy if you're working with more technical parts. So for an engine, for instance, that you can really, you can build the digital twin of an engine and then you can build the physical engine and then you can test and you can have it, you know, understand. But when you're working with cities, there's so many different things that. Where you can lose trail because this is so complex, so it's hard to understand if we really are getting the results of course that is coherent with the real world so, but I think that's a very interesting area of innovation to how can we validate and if we are talking about the representation In the Digital Twins, it's of course if you are gathering information about people, sometimes you end up with information that could be sensitive. But I think also it's very important to use data that really shows the pros and cons with the city as well. So there are many different kinds and ways to collect data and make it anonymous. So, so you cannot trace this. If you have, if you are creating, if you like our avatars, they weren't actually real people, they were they just behaved as we have like invented. But if you gather a lot of data from real people, you can transform that into being more anonymous that they are. They are behaving like the real people but they are not real people. So so that's that's important if you're working with the integrity questions that and that you that you that you filter. Maybe the information that could be sensitive but but it's hard because you put a lot of information on top of each other. So if you look into one information set maybe that's totally OK. But if you combined it with that and that and that maybe then you can see your pattern or or something that that could be sensible to the sensitive to to show so. I think it's.	DC DSU DS	ED MD TCA
00:42:05 SPK_1		

Yeah, and you mentioned equality as well.		
00:42:10 SPK_2		
Yes, I think equality is. There are so many questions about that in the digital twin. Oh, I see my. I don't have so much battery. I switch over to my phone if my computer and the the quality, I think it's both the way that you visualize it. How could you visualize the quality? It's very easy to visualize a car or a building or a tree or a fence or whatever. It's very easy to just put in there. But if you want to visualize segregation, or if you want to visualize how safe do I feel on this spot? Or how can if I have disabilities, can the environment help me if I'm if I have both physical And what do you call this phykiska?		
00:43:23 SPK_1		
Mental.		
00:43:25 SPK_2		
Mental disabilities how how can how can we visualize that and I think because that's so much harder so it's very easy to to be to visualize a building but if you want to visualize how safe does this area feels in different times of day and. In the middle of the day, this is maybe a very nice place to visit. But at 3 o'clock the night between Saturday and Sunday, maybe you feel very it's not in a safe place to be. And how can you visualize how the city is changing over time and and day and how people feel about it? Because I think that's important when you're planning your city to and. To to and and they work with it of course they do work with the question but the thing is how can you bring it into the the digital twin. I think it's that's very interesting to to to find ways and then you have the equality and who is decided what to put in the digital twin. Where do the data come from? Is the data biased in the 1st place? And how can you is could you, by using the digital twin, find bias data? Because when you are starting to visualize things, you often understand that. Oh My God, let's see, we don't have any information about this. We really need this but we don't have the information or the information is so poor quality or whatever. So so the digital twin and the visualization can help you like reveal the the the things that you should have put in there but that is missing. So I I think it's a good way of of having the digital twin as a checkpoint in some way for a lot of different questions.	DQ DC DSU BD	MD ED
00:45:45 SPK_1		
Yeah. And this is being actively discussed in the project or what's the phase on these kind of questions?		
00:45:55 SPK_2		
Yes, we are discussing it every day and we do did it a lot when we created the <Platform>, but but I cannot say how much of this discussion that is going on. From the other partners, we were <Number> partners. So I don't know how many of the other partners that are still discussing this, but we are because I think it's very important and I think it's a lot of interest around the social questions as well.		TCA

00:46:26 SPK_1		
So talking about bias.		
00:46:27 SPK_2		
Because they are a part of the, They are a part of the Agenda 2030. So you need to take care of the the questions surround not only the, yeah, all the goals you need to take care of, yeah.		
00:46:46 SPK_1		
So talking about bias, to what extent do you think that certain bias can be acceptable when you are doing the project and using digital twins to reaching certain goals?		
00:47:02 SPK_2		
I think you're always using bias data in some way because they've been working with the data before the digital twin or the digital twin is created on the data that they are using. So if the data is incorrect or biased or not enough quality or whatever and that doesn't change. By making it into a digital twin, the data is the data. If you have poor data, you have poor data. If you're working with it directly in one, in one program or just a text or whatever, or if you put it in a digital twin, I think. But one thing I think that the digital twin really opens up to is a discussion about data quality. And I and I think that's when you're starting, we're working with digital twin you you always starts to talk about the data and just starting talking about the data. Is this good enough for what purpose? How can we put meta data on this so people understand what is this kind of data and who is responsible for this data, how often is it, is it updated, how? So I think the digital twin, the discussions around the digital twin can really help the city or some other people working with information to start discussing information and have the knowledge about the importance of understanding the information because you can use very good, very good data and you can use very poor data and they can all be a part of the digital twin the the tricky thing is to understand which is which. So you know that okay this bridge for instance is a 3D model from old photographs or something. It's very poor quality. But this bridge is like top notch. You have the perfect everything is is correct as long as you want to go. So you need to be able to to see and understand that in the model. But I think really I have been so many times when I work with 3D models of cities. You you you see things in the data. When you start to visualize it, you see things that is this really correct. It seems like no this this could be right. And then you start to look at the deeper and then you see that okay this data was wasn't good enough for all. So I think just to have anything in front of you in the way that the 3D model does. It's very good for vi atta avslöja reveal or when you have a biased data. So I think it's it's like a good tool for understanding the data that you have. So, yeah,	BD DC DSU DS DQ	TCA
00:50:26 SPK_1		
Thanks. Did you have another? Oh yeah, talking back about the real time data, so after collecting data. Who will be the one that who owns the data having the ultimate ownership of the data, yeah.		

00:50:44 SPK_2		
That's that's hard. We talk about this a lot because. Some people are collecting or some companies are collecting a lot of data and there was a slogan or like a sentence that a couple of years ago everybody said that the data is the new oil or new gold because there was something in that, that you can make a really good profit or just selling data like on a stock market or something, I think. I think information, good quality of information is always it has a value and of course if we have somebody's collecting data of very high value, it's well defined, you can understand the data of course it would be interesting to to buy or license for use or something like that. So, so I think the. The one that is collecting the data of course is is the owner. But I think I think there will be a lot of more openness open data around this as well that for instance, people sharing their data freely like when you click in those in your social media you say yeah okay to to to share this data if I have this or that or. So I think I think this with real time data is is we are going to see a lot of different solutions I think but I don't maybe not, yeah I don't know if it's the new gold or new oil or not, but I I think it's important as well that that you have the the time frame that if you're collecting data can you see over can you see it over time or because you yeah and and start to understand and and see patterns in data. I think that is very interesting and and and if we can see that I'm. I'm a fan of of of openness and sharing. So. So I think that if if I have a data set that I'm very proud of and both of you have other data sets and we can sit together and say, oh, what if we combined all these three, that would be awesome and then we quit Stop talking about selling this to you or not to. But we we three together, we can see that. At least it's going to be great if we do this together and then we put in our data and we do this together and we don't discuss if your data is more valuable than mine and so on. That is what I think could be a fantastic way of of getting ahead quickly because it's the benefit for all and so on kind of way I think.	DC DSU DS	MD ED PD
00:54:07 SPK_1		
Yeah, I'm sorry. I see now that the. The time is 12. I was just wondering do you have more time or?		
00:54:14 SPK_2		
Yes, I have. So a couple more minutes, Yeah.		
00:54:17 SPK_1		
Just let us know when you don't have time anymore than I want to ask about Yeah, yeah. So for this then collecting data about citizens and their behavior, do you think there's any way that the Digital Twins and the decisions made based on it might affect the ability of Citizens to choose for themselves, if that makes sense.		
00:54:48 SPK_2		
If to choose if they want to be visible in the twin or not or how do you?		
00:54:56 SPK_1		

You I'm thinking their behavior is like.		
00:55:00 SPK_1		
Should I? Yeah. I can supplement something because I remember one of the use case that you mentioned is that when there is heavy rains and then the traffic and then how people's move to which routes they are choosing. So it seems like that for digital twins is already have some designated routes and simulate those solution for them for their choice. So we are thinking if actually the design of digital twins is actually changing people's behavior. Then some kind of things and idea towards that. So what do you think about it?		
00:55:33_SPK_2		
Yeah, I don't know if I understand. Did you mean that if we're using the digital twins to and we look at people's behavior and then we build things to meet those behaviors that we can, if we are like building something that we we want them to do.		
Timestamp_SPK_1		
Yeah exactly. Yeah. Yes.		
00:55:58_SPK_2		
But I think, yeah you could like nudging if you're it's always if you if you make something easier than another thing so it's easier to do the right thing maybe from we say sustainability. If we have more a choice that is more right for sustainability than another, like the reducing your car and using your bike instead, if you want people to make that change, that's about like nudging people over to more sustainable behavior. So so of course. A very lot of functions are trying to nudge people because you cannot, you cannot, you cannot tell people that you have to do this. But you can inspire them. You can nudge them to do what you want. So of course you can use the digital twin as well as we talk about gamification or we talk about. Trying to to to have different behaviors so so of course you can you can try as an example if you have different alternatives for a city area and then you can try the different alternatives on the avatars, I see which alternative of creating this area will make. Most people yeah and do do different things so. So I think that is important to to try to simulate to have the the best city we can possibly have and if that means that you want people to. Maybe go by bus instead of car because you don't want to build a lot of broad roads through the city. That could be a thing that you can try but I think and you can inspire. But I think you can never yeah, I think you you cannot change people to to all the The very different levels, but you can inspire them. And we're working actually with another project called <Project Information>. How can we design things for the sustainable society? Because we haven't. There's a lot of things that is decided today and designed today that is not leading people in in the sustainable direction, so.	DSU	
00:59:11 SPK_1		
Is that some kind of like guidelines or what's? What's that project about the?		

00:59:18 SPK_2		
I know we are. We are trying to, we are working with the young adults \<Project Information> and we are and people that want to make sustainable decisions and we are talking to them about what do you need. And to make a sustainable decision, can we create some kind of digital tool or digital method or digital thing with virtual reality when you that you can use to inspire yourself or to learn things that leads into a more sustainable behavior? As young adults, if you're doing the right things when you're young, like yeah, there's a lot of things we are looking at. Hopefully you keep those habits even if you're through your lifetime. So there's a lot of things that you can. Innovate with this new technology with VR or visualization. I'm not only working with digital twin, that's just a part. So this is more of virtual reality project, so it's interesting everything with behavior. We are people. We are not computers. We have people and we have our behaviors and. ==And we want people to have equal and healthy lives.== So how can we work with this technology to democracy questions is so important with you when you're working with these technologies? ==I think can we assure that we have a sustainable democracy as well as sustainable environment so. And I think it's very interesting when you when you take the digital twin out in the city and really and really enjoy you working with them together with other people.== I know that the \<Location Information> has been doing \<Project Information> projects in different areas of \<Location Information> to have schools to to start to discuss the city and the environment, and they go into \<Project Information> and have opinions about their own area and so on, try to believe things and so on. ==And I think that's important from the democracy perspective as well.==		MD ED PD
01:02:12 SPK_1		
Is that something you plan to include in the digital twin as well the discussion with citizens and getting feedback and opinions.		
01:02:25 SPK_2		
Yes, we hope to in a more more extended way than we are doing today. We we are working with a project that we want to engage people in, have some discussions with where you can have local farming, how, how could that look like and how can we discuss that question. So there's a lot of questions with how do you develop your city together with the citizens and that's important.		
01:03:12 SPK_1		
I don't have any questions, no. Yeah, no, I was this is a bit of a topic from what we talked about just now. But I was thinking when you when you take data from that was collected for one purpose and transfer it into a system that has another purpose, are there any considerations with like goals and yeah, maybe bias or how can that? How has that worked with?		
01:03:43 SPK_2		
When you transform.		

01:03:47 SPK_1		
Like data that is collected for one purpose and you put it into another system with another purpose.		
01:03:51 SPK_2		
Yeah, yeah, of course it could be. It could be if you have gathered information with a special purpose and special time frame or Whatever, It's very important that you understand the data set and when you're starting to use it together with other data sets or for another reason so so I think the the quality of the the metadata around the information set is very important because everyone must be able to understand and make your own decision about this data set, Is it good enough? Is that the right data? So of course it's always a bit tricky when you start to to to change to move data. If you don't know what you are doing. It could it could end up being wrong data at the wrong place but but I think. No, I I I don't think well I I I think people, I think people know and understand people that are working with information that they understand that you should be very careful when you when you're making consumptions from something that is not that from the beginning. If you are trying and if you are sketching and if you are testing, it's OK, but if you are doing something in a very scientific way. You really do something that is has to be correct and scientifically correct. Then you then you really need to understand. But for just playing around you can do like a little bit. You're more free when you are testing because then if you're testing then you always go back and do things in a proper way when you have the results from your tests. So I think.	DQ DSU DS DC	
01:06:08 SPK_1		
So you try it in different scales?		
01:06:12 SPK_2		
Exactly. Yeah, exactly.		
01:06:13 SPK_1		
Okay. Thanks.		
01:06:14 SPK_2		
When you're scaling up and when you're doing things that really affects people, then you need to have everything in order. But when you're trying to learn how to work with new technologies, it's more of than the main focus is to understand the new technology. And work with the new technology. So if you don't have everything correct, like in our avatars, the question wasn't if the avatar was working exactly the right way or not. The question was, is this a way that you can start talking about these questions? So then the purpose wasn't. Really, to analyze exactly the avatar, it was like an illustration or a picture, just to illustrate or to demonstrate something that you could have if you're doing. If you're developing a tool that can handle this, then you need to have the right information, of course, and the proper data. But just	DSU	ED MD

to talk about the usability of the digital twin, then you can try with data that is not that accurate. Yeah. Mm HM, because it's a demo or prototype. Yeah. Yeah.		
01:07:36 SPK_1		
Okay. So are there any further limitations or concerns that you see that we didn't discuss yet so far?		
01:07:50 SPK_2		
Yeah, probably. Probably I thought.		
01:07:51 SPK_1		
Yeah.		
01:07:53 SPK_2		
That's all. Okay now my computer. It wants me to shut up. Maybe I just. Go switch to my phone so I won't lose you. Let's see, I couldn't go in there, so let me just let me just have my cable.		
01:08:32 SPK_1		
OK. Thanks.		
01:09:32 SPK_2		
We're keeping it alive. For a couple of minutes I'm Sorry for that.		
01:09:38 SPK_1		
It's nice. OK. So any limitations or concerns that we didn't discuss yet that you would like to discuss?		
01:09:57 SPK_2		
I don't know, maybe something pops up, but something I I think it's. Something that I think it's interesting it's a lot of regulations from from Europe. I I don't I'm not involved in all these things. I try I have a colleague that is is trying to understand how that could affect the digital twins but I. I think we don't know yet so much about I I don't know yet, but I really want to know more and that's why I thought your work was very interesting that there are regulations coming from Europe and. And I think it's important that that we understand how would this affect the work that we are doing and what new regulations are we interested in in Sweden to to change or create or what do we need to get forward. And I think the regulation part is I think it's a very important like frame or to to to have to have a framework that you can relate to. I think it's very important and to be able to understand what's coming from, from Europe, what's happening out there, because when you're working, you're in your little own project and and everything, it's hard to to to understand what out there to to. Keep updated, keep yourself updated and what's happening you don't have really the the time really you know that something is happening. There are a lot of data acts and so on discussed in in in Europe and and		

it's hard to to understand but because when you're trying to understand it you you realize that. You don't because it's very hard I think to see the effect. How would it affect us? Yeah, I think of course it will affect us in in different ways, but I think it's hard to understand how and is there anything we should do to like try to convince or? Påverka. What's that?		
01:12:39 SPK_1		
Effect or influence.		
01:12:42 SPK_2		
Influence our own government. This is important for for us working with those things and I think the the national hub for digital twin, they could really have an important role in trying to to inform the decision makers that. These questions are important and this will affect us in this way and and and I think everything now is like up in the air, everything is so fussy and you hear something there and you hear something there and I think it's very hard to to to create a picture that you can relate to. So I really think that would be a benefit for for working with digital twins if I could know more about regulations. So it would be very interesting to see what you're going to find in your work. And I think when you're working, I'm sure you will pop up thousands of ideas or thousands of questions that you cannot. That you cannot have in your work because I I understand that you are limited as well, but it will be interesting to have those questions, the questions that are like leftovers from your project. And see is there anything important here that you had to put aside. But it would be important to to go further in another another context. Because I think you the questions are so, so important and it will be very nice to to see what you can do from all these talking that I've been doing and and I get, they've been talking to a lot of people and they have all different angles and all different things that they think is important. So because we are all working on different layers. And different levels and we'd like going forward like everything is doing very little piece. So I think it's very interesting when people try to understand more regulation parts of it.		
01:15:20 SPK_1		
Okay. So I think that's yeah. Any questions, do you have any questions for us or anything else you would like to discuss?		
01:15:29 SPK_2		
No, I just have one question. Is that the material that you're using? I read your, I got a letter e-mail from you with.		
01:15:47 SPK_1		
The content The consent form.		
01:15:50 SPK_2		
Yes, yeah, consent form. But the things are are you, how are you going to use the, the video material? Are you going to use it just to take out parts or are you going to		

use it for making quotations or how are you going to to use the material or is it going to be, yeah, anonymous or? So.		
01:16:19 SPK_1		
Yeah, it's going to be anonymous and we're going to transcribe the interview. So we have everything in text. So the videos will not be part of any anything public. That's just for us And the material will be I think summarized like from the different interviews, the parts that are from the same theme or. Will be put together and explained and there might be some quotations, I guess, if there's anything that's very important that we can't rephrase without losing value to it. So yeah, that's. Did that answer your question?		
01:17:07 SPK_2		
Yeah Yes. That is and so if you have any other questions or if you. Didn't if you want to have any following questions when you're looking at the material then you are welcome to just call me again or also ao and I really want you to invite me or to send me your material and so on so I can see the whole picture when you are done. So when are you going to finish your work?		
01:17:43 SPK_1		
Hopefully we will finish everything in June.		
01:17:46 SPK_2		
In June, yeah. So you have something to do with them, Yeah.		
01:17:51 SPK_1		
That would be great. If we got any follow up question then we will e-mail you to. Yeah, yeah.		
01:17:58 SPK_2		
Yeah, Yes, good. So, so, OK.		
01:18:02 SPK_1		
That will be all for today. Yeah. Thank you. Thank you so much for your time and.		
01:18:07 SPK_2		
Thank you very much and good luck with your work.		
01:18:11 SPK_1		
Thank you and have a nice day ahead.		
01:18:14 SPK_2		

Yeah. Thank you. Bye, bye.		

Appendix 7: Interview 5, P5

Theme	Colour Codes	Factor	Text Codes
Context	<mark>Yellow</mark>	/	/
Data Management	<mark>Blue</mark>	Data Collection	DC
		Data Storage & Use	DSU
		Data Sharing	DS
		Data Quality	DQ
Algorithmic Bias	<mark>Red</mark>	Data	BD
		People	BP
Principle of Discourse - Democracy Principle	/	Pragmatic Domain	PD
		Ethical Domain	ED
		Moral Domain	MD
		The Theory of Communicative Action	TCA

Participants	
Authors	SPK_1
P5	SPK_2

Transcript	2nd layer	3rd layer
00:00:00 SPK_1		
Thanks. Recording in progress. OK, Yeah. OK. So do you want to start by telling us a bit about your role right now, day-to-day?		
00:00:17 SPK_2		
Yes, my role is that <Role> at what they call it, <Department> In <Location Information>. <Company Information>. But I'm the <Role>. Yeah.		
00:00:51 SPK_1		
Okay. And would you tell us a bit more about the Digital twin or twins that you're working with? In the city setting.		
00:00:59 SPK_2		

Yes, I think I have not. So I have not have so much knowledge about the digital twin because there are other people at my job that is more have a better knowledge about the digital twin, but my knowledge is more about traffic analysis and so on in the digital twin. So, but I can say that the digital twin is we have worked with the digital twin for several years now and yeah, it's getting better and better. <Location Information>, so. Yeah, that's what I can say. I can send you some information on of other people that can tell you more about the digital twin. <Name> is a very good people. Sorry some.		
00:02:13 SPK_1		
It's Okay.		
00:02:18 SPK_2		
Sorry, sorry. Sorry. That's my cold, that is.		
00:02:25 SPK_1		
Yeah.		
00:02:25 SPK_2		
Yeah, but the <Name> is a good person to talk. Talk with if you want more specific details about the digital twin.		
00:02:38 SPK_1		
But you're working with a digital twin, so that's that's also interesting too. Are you working in it or what's the?		
00:02:47 SPK_2		
No, I don't work in it. I work with the digital twin as an instrument to how we can use the digital twin in the into traffic analysis areas so. I have not the knowledge about the digital Twin, but more about how we can use the Digital Twin in that area. Yeah, Okay, yeah.		PD
00:03:18 SPK_1		
Yeah, So what are the?		
00:03:20 SPK_2		
I can recommend you to talk to <Name> of if you want more details about digital Twin, is really, really good in. Yeah, our digital twin.		
00:03:33 SPK_1		
Maybe you could send his email?		
00:03:35 SPK_2		

E-mail, yeah, yeah.		
00:03:37 SPK_1		
So what are the organizations and companies and people that are involved in this project, like the stakeholders?		
00:03:50 SPK_2		
In that project that I think that you referred to was the <Company> project, Was it? Yeah. And the stakeholders in that one was <Company> and <Company> and it was <Company> financed project. Sorry. And and what can I say and the stakeholders was that it was a few consults companies that were. Have a good knowledge about the. Sorry, I'm so sorry.		
00:04:45 SPK_1		
It's okay. Just let us know if you're not. If it's not working for you, it's okay for us.		
00:04:53 SPK_2		
If it's okay for you, it's okay. Yeah, it's okay for my yeah, yeah, sometimes, yeah. And what can I say, we have some consultants companies that were having good knowledge about developing a AI products and and also digital twins and so on that we were working with in that project and the purpose of the project was to see how we can use the digital twin more. To say this decision making in the about the traffic solutions and and our focus was on the pedestrians and so on that was walking in the city and we want to explore about. The behavior, the behavior of the the walkers in in cities, it's very hard to what do you say predict how, how, how they will walk and so on. And therefore we want to explore about how we can use to use use the digital twin to see how the walkers are walking in the city, and in this case we try to have some function. For example we have. You can put in some rain in the digital twin and and in <Location Information>, so it's a very good function. And when we put in rain in the digital function and then we see how the avatars that we have in the digital twin, in this case was walking when it was raining. We also tried to change the length of the fences and so on between the public transport areas. And see how the the walking flows was that they they are how they were walking with fences and so on. So yeah it was more about how to understand and explore the digital twin, how we can use it for more decision making in the in the development, the urban development, yeah.		
00:07:50 SPK_1		
It's very interesting to hear the ideas of using Digital Twins like the use cases. But I want to understand more about how do you guys comes up those ideas like what to do with the digital twins. Like you mentioned about that when there is weather changes you want to use it to predict like how the pedestrian they will walk or behave. But how do you like comes up all the requirements is it come from something that you did some research before or something like that? We want to understand what's the story behind those requirements.		

00:08:22 SPK_2		
Yeah, the story about the requirements is that the walkers in the city are getting more and more focus in the urban development and also bicycles and so on and also public transport, but, as we can see now that we are getting more and more bicycles in in the <Location Information>, which is very, very positive, we also get more and more walkers and we have a conflict between the the bicycles and walkers and we also want to change the focus from car driving more to walkers and and the bicycles and and so on. Therefore we therefore the walkers are very, very, what do you say? very, very important for us just to understand. And therefore it's very good to have some tools just to ==to use in the in the urban development to to see how it we can build a good city== for the walkers and we also have, we also have some functions regarding to the regards to the workers. We have ==Traffic Safety== as a very prior issue and also how. What do you say? We are getting good, We need to have good. So I forgot the word.		ED
00:10:15 SPK_1		
Try Swedish. Yeah, she can translate.		
00:10:19 SPK_2		
Stadsmiljöer.		
00:10:21 SPK_1		
Like the city environment or.		
00:10:24 SPK_2		
Yeah, utformnings beslutsfattande kring		
00:10:33 SPK_1		
A decision making around shaping the environment, the city environment.		
00:10:37 SPK_2		
Yeah. It's shaping environment. Yeah, exactly. Yeah. Yeah. Thank you.		
00:10:41 SPK_1		
Is there any citizen that is actually involved in the discussion when you comes up those requirement?		
00:10:49 SPK_2		
Yeah, absolutely. As a in the <Department> it's we have a lots of function. It's a. It's a conflict between a lots or lots of functions. It's the how the big areas are for the bicycle areas, The Walking areas, the the trees, the the public transport, the cars and and we have a lot of people in in in the development of the environment and so on with different requirements and. Therefore, it's very, very good to have this access to a digital twin just to you can sit down between lots of functions and you have a		TCA MD ED

talk about how it how it is, today we are have some papers and sketches and so on. You don't get into 3D world so. So it's very very good to use the digital twin just to have a meeting around it and and click and check and and adjust and so on just to to get a good solution. Because every function wants to need wants to have a very very. A big area for your function, if you want several, it could be trees, it could be, yeah, the public transport and so on. But we don't have that space today, so it's a it's a fight thing about the space in the, in the urban development phase. So therefore the digital twin is a very, very good tool to come together and in a world, yeah. And yeah. And also Traffic Safety, yeah. Because the, the, the fences can be in an environment point of view the fences could be not so, not so good, but the fence is very good for the Traffic Safety also it's lots of function that will you have to work together with the yeah.		PD
00:13:37 SPK_1		
Yeah. So you use the digital twin to reach some kind of consensus or how? How do you?		
00:13:43 SPK_2		
Yeah, exactly. Yeah, I think so. I think that that was the best results from our project in the <Project Information> and we are not there yet. We have a very, very long way to to take us to what do you say working process to use a digital twin in our daily work. But the best results from the from our project was that it it's very good to sit down around the digital twin just to to between some functions and also to describe the the traffic flows when you, when you talk to a traffic analyst figure, It says very, very lot for a traffic analyst. They understand it quite the way that if you if you got to figure in in for a specific road or for walkers and bicycles, they can use that figures to understand how it will work in in the real life. But if you're not a traffic analyst and you want to see how. How the traffic called situation will be in the in the future. The digital twin can be a tool to look at for some decision making. This decision functions just to understand how it will be. So yeah if you understand what I mean so yeah.		PD TCA
00:15:35 SPK_1		
But when you are using digital twins to like discuss or make decision making and coming to a consensus, there is always people have different agenda or conflict of interest. Then how can like who will be the one who make the final decision or by what consideration or criteria to make the final decision.		
00:15:55 SPK_2		
Yeah, it's so it's very, very good question. It's depending from time to time. I cannot answer it. Yeah, it's a very, very good question.		
00:16:10 SPK_1		
Yeah, I know it's very difficult to answer as well.		
00:16:13 SPK_2		

Yeah, yeah. It's very, very difficult because it's all depending about the area and the situation. And and yeah, it's a very good question, but I cannot answer it because it's it's. <u>It's all about the situation and it's depending on the which interests</u> that are <u>more critical and it could be</u>, yeah.		
00:16:52 SPK_1		
OK. So is there like how do you rank what's more critical is like safety I guess is? Critical Do you have any examples on criterias that are more critical than others?		
00:17:10 SPK_2		
Yeah, absolutely. <u>Safety</u> is very big, but it's it's depending about the situation. Sorry for my not so specific answers in that one, but yeah. <u>It's all about specific areas in the situation,</u> can we? Yeah.		ED MD
00:17:40 SPK_1		
So our thinking talking about policy, is there anything that policy making is actually affecting the development of the digital twins or the projects?		
00:17:53 SPK_2		
Can you take the question once more? So I understand.		
00:17:56 SPK_1		
For the how is the policy making is actually affecting like implementing the digital twins or how you shape the projects or any requirements that related to digital twins?		
00:18:10 SPK_2		
Right now we we test the digital twin in this project <Project Information> we also tested it in one project that we put in some figures from the traffic the analysis that we did in some projects to see how the cars was driving in the digital world and so right now we have it's very good to to have that question also to <Name> also about the digital twin but. <mark>The the focus has to right now is to build the digital twin and how we will work with it and how we will use it.</mark> We have little bit working to do in that area because to therefore we have tested in the in the project that you found about the central station we also tested it some other different projects but there we are right now I think the the coming years the focus will be more on how to use the digital twin not to build the digital twin if you understand what I mean and in that case you have to see how the. <u>How we can use the what do you say about the policies and so on? What's what's how we can use the digital twin in the best way about the working process and so on and also about the the policies and so on.</u> Yeah, but maybe <Name> can answer it better than I can. But from my point of view, I see that we have. There's been lots of focus on the building the digital twin. We have started with the focus focus to working in the digital twin, but it will be work more focus on that. The coming years something.		PD
00:20:40 SPK_1		

Maybe we can go for a broad question. Or should we? What do you think? I was thinking? OK, yeah. Yeah, this is a a broad question, but in your opinion, what do you think is the most important to think about when designing and implementing a digital twin or system in city planning like in the city context?		
00:21:07 SPK_2		
I I think, I think the competence of the digital twin because right now we have a lots of. What you can say old, competent, old competence in in city planning we have experts in Traffic Safety, we have expert in the urban planning, we have experts in the environment and so on. But they don't have the. The competence in the digital twin, we have also the the competence in the digital twin, but we if we are setting up an organization or working process that we must use the digital twin more in our project, ==we need to have a lot more competence in the digital twin area because the the experts cannot use the digital twin the in the same manner that the experts are in the product.== In Central Station we have some experts that are sitting down and developing the the avatars and so on the the fences. ==During the conversation between different experts in the in the urban planning, but it we don't have a working process for that yet in the urban planning process.== So that's the focus in the future I think just to to to have the right balance between the digital twin competence. We have the digital twin competence now, but not in the not as, not so much as we have in the experts. If you add the products, if you understand. So I think the digital twin is. Could set up a new organization for urban planning in the in the future. If you understand what I mean.		
00:23:53 SPK_1		
So actually I'm interested in your viewpoint, like to what extent do you feel confident about the technology that used, like Digital Twins? And also how about the companies that is actually delivering the Digital Twins for you, achieving the goals? To what extent do you trust in that or like the confidence?		
00:24:17 SPK_2		
Yes, if I understand your question right here we have a conflict between and today I have the traffic and this perspective that question because that's my, that's my work but we have some when you do a traffic analysis today, we have some programs today that will get you a figure. We also have the <Company> software that we can also see some. Yeah, sorry. In the in the <Company> software we also can see 3D dynamic video, but not so good as in the digital twin. It's far away from the digital twin that we are setting up. But that there is a conflict between the the, the older way of using traffic analysis, the software and so on and the new area, and we have to explore that used to see how we can use these tools together because the we need a digital twin in the future just to explain how the city will be for them that are not the traffic analysts just to see how the cars will how it will look at a place in the city about cars and bicycles and so on. And yeah, so yeah, ==I think it's it could be a conflict between old traditional methods, new methods,== yeah. Hopefully I answered your question.		
00:26:41 SPK_1		

Yeah thanks. And also I heard that when we asked about the participant in the projects, you mentioned about some consultants, it's also selected in the projects at first and then so is there any criteria that or consideration when you're choosing those consultants like based on what things that you chose them for this projects?		
00:27:07 SPK_2		
These consultants it was a <Company> project in this case. So the consultants was I think every company that was in this project <Project Information>, mostly of the project that it's self financed if you understand what I mean and in that case the consult companies see these projects as a good experience to build some competence and get the new ideas and so on. Yeah, and yeah, so the company is that was and I think it was very, very good because in sometimes we get companies and the consultant companies that are very interested in this part <Project Information>. Yeah. Sorry, sorry.		
00:28:59 SPK_1		
Yeah, that's fine.		
00:29:00 SPK_2		
So they are very, yeah, interested in the project. Therefore they will oversee that they do not get 100% payment back, <Project Information>. Yeah, you understand.		
00:29:18 SPK_1		
So Okay, so it's the companies that are participating are doing it out of their own incentives.		
00:29:28 SPK_2		
Yeah, I think so, because they don't get 100%.		
00:29:31 SPK_1		
Yeah. MM HM yeah. And.		
00:29:35 SPK_2		
They're very interested in these types of questions. So it could be very, very good because so yeah.		
00:29:42 SPK_1		
Okay. So shall we go to the other question or you want to do that? Yeah, we can so. For the digital twins that for example used in the weather change use case or the pedestrian, how they move when there is different change of weather for those data, do you know that it's actually a real time data provided by the traffic department or actually the data is collected by other companies?		
00:30:18 SPK_2		

It's it's depending and I think in this case we used the traffic data from our but it's it it can depend sometimes and and the flow the traffic flows for pedestrians or cars so on we we can use it for from our own old methods, the figures, into the digital twins so that's that is our figures that we use in the digital twin, yeah.	DSU DC	
00:30:56 SPK_1		
For the data, does this involve any like sensitive data like people's data or it's just cars, trains, traffic data?		
00:31:07 SPK_2		
Can you say it? What do you mean?		
00:31:08 SPK_1		
For the data that you mentioned, it like from a extracted from the <Department> Does it contains any personal data like sensitive data or it's just plain cars?		
00:31:19 SPK_2		
Yeah, yeah.		
00:31:20 SPK_1		
OK.		
00:31:21 SPK_2		
We don't use any personal information in the it's only it's I think we have if we go with cars, we only have private cars. We also have what do you say more heavier vehicles. So we have we have three different types of traffic. Private cars, small cars what's that what more what do you say lighter vehicles and heavier vehicles so we have three different categories so it's no personal data because we we have we have a problem in our models is to just to see where the car goes from one point to another point, because we cannot collect the data the the station numbers and so on. So we have to modeling the data, yeah.	DC DSU DQ	PD
00:32:35 SPK_1		
OK. And that's because of GDPR that you can't?		
00:32:40 SPK_2		
Yeah, exactly. Exactly. We can't do it, but it's we. It's lots of work just to establish or organization to take care of that type of data.	DC DSU	PD MD ED
00:32:55 SPK_1		

Yeah, so.		
00:32:58 SPK_2		
It's possible, but yeah.		
00:33:02 SPK_1		
So to anonymize the data, it's being a bit simplified, would you say or what's?		
00:33:09 SPK_2		
Yeah, I think it's the best way also. Mostly of the analysis, we're not interested in personal data. We are only interested if it's a small car or heavier car that are or go get, go get the street. So in most cases we don't need more detailed information than that, but in some cases we want to have more data from point to point. So in that case it it's it could be to have more personal data or more detail data. But yeah, if if if also have detail data, it's. It also depends on depending. If you if you don't have any data but registration number, so on we we if we if you already know exactly where the point is, where the car is arriving or leaving from a point it could you can understand what person it is that have traveled that way. So in that case we have most aggregated data at a very, very higher level also. So yeah. Okay. So we are working in bigger areas, yeah, mm HM.	DC DSU DQ	PD
00:34:51 SPK_1		
So for the data that have been using in the Digital Twins, to what extent do you think that is actually showcasing how the real world is working? Because some of the data like is simplified, so when you put it in the digital tool, do you think that something is missing when it's looking bad?		
00:35:18 SPK_2		
Yeah, but absolutely it could be more about the personal things in the digital twin. Absolutely more about the personal things we're not as you see in our picture we have avatars that we it was white avatars and absolutely you can you don't have you don't have the detailed things in the digital twin about red cars or something that that could be for the. Yeah, For what would you say for the oh, känslan I rummet, om du förstår, oh, can you explain that?	DQ	
00:36:20 SPK_1		
The the feeling in the room. Yeah, yeah, yeah.		
00:36:23 SPK_2		
Yeah, yeah.		
00:36:25 SPK_1		
OK.		

00:36:26 SPK_2		
I got it, got it. If you look at the. Yeah, if you. Yeah.		
00:36:35 SPK_1		
It's interesting.		
00:36:36 SPK_2		
So, yeah, so we we in some cases we don't have any more detail that kept, but I think it in a total we have so much other important things in the digital twin that is more necessary and the more details about the cars and the personal information about the walkers and so on. So I think it's debatable ehm, yeah, things to have, yeah.	DC	
00:37:16 SPK_1		
Okay, so we wanted to ask you then going into the next theme here. What's your current understanding of algorithmic bias?		
00:37:30 SPK_2		
My whole thing Yeah, that's really low. Thinking about the like, yeah developer the I don't have a I'm working more on how to use the digital twin if I understand your question the right way.		
00:37:45 SPK_1		
I'm asking what's your like perception of algorithmic bias? Do you? Have you heard of that before?		
00:37:53 SPK_2		
No		ED MD
00:37:58 SPK_1		
So that's to briefly explain, there could be when you collect data and do algorithmic processes like algorithmic calculation and AI, there could be differences in the data that maybe you have more data on a certain point which makes the prediction about that point more accurate and maybe the data on the minority then point would maybe not be represented correctly. Or so there could be outcomes that affect some people negatively because the data is skewed somehow and that could also be In the algorithmic calculation or from the people building the algorithm, so there's different types of bias there. Is that something?		
00:38:56 SPK_2		
That I understand, yeah, then I understand what you mean. In my department we are have some expert on on that one, but not the algorithm building, more about	DC	ED

the method used to get. How do you think to have the right answers of different parts of <Location Information> and so on. So we have a a thought about how to how, not how to build an algorithm in the the digital twin but more about the the statistical. Sorry for for methods about. Yeah, I think it's a better question for for <Name> in that case.	BD	
00:39:44 SPK_1		
OK, OK. OK. So so this bias concept, it's not something that you discuss in the project in general, is it more for the technical people?		
00:39:57 SPK_2		
I think it's more technical that we have some methods about that. We need to have what you say, as you described, we have, we have to make sure that we have a correct population from each area in <Location Information> and so on. And our knowledge in our group is more about the methods about that, not how to develop These methods in the digital twin, these algorithms in the digital twin in that case we don't have any knowledge about that but we have the methods about population and so on. But I think you can also unless you can ask <Name> about that it it.	DC BD	ED
00:41:02 SPK_1		
Yeah, Okay. So when you collect data, you have some kind of requirements for how it should be distributed or yeah.		
00:41:17 SPK_2		
Yeah, yeah, we have. But more on not in the digital twin	DC	
00:41:24 SPK_1		
In the statistical sense?	DC	PD
00:41:26 SPK_2		
Yeah, exactly.		
00:41:27 SPK_1		
Okay. Is it a industry standard or just a guideline that from the traffic department?		
00:41:35 SPK_2		
Yes, in the, yeah, in some areas we have that, but it's more about our our knowledge and methods to get it correct. But if maybe the few few areas we have for yeah overall guidelines.	DC	ED MD
00:41:59 SPK_1		
Okay, yeah, I yeah, actually I'm interested in, I think I also asked this question to a lot of people and then it seems like a very difficult question to everyone because for the		

one just like the one example like. When there is weather change and then there is a lot of simulation and then in the simulation is trying to like find out the best option for the pedestrian or best option for the cars traffic and things. So do you think that to some extent is actually limiting peoples of their choice of their movement? Because for the digital twins they simulate different kind of solution so it's limited to several solution only. But actually, maybe human being, they want more choices and want more options. So do you think that actually it's limiting people's choices on that for your opinion, personal opinion?		
00:43:07 SPK_2		
We it, it, it is product. But the station and the avatars it was we we talked a lot about how the behavior of the pedestrian, yeah. And it's very very complicated to to to describe the the behavior from pedestrians. So I don't remember exactly what we what we landed in in that that discussion. But we discussed it very very much about how to how to set up in the the algorithm how our avatar will behave because we have so many different behaviors for a pedestrian either environments but maybe I think in the future the AI products we can use more about the behavior that we if you have a if you can measure an area with a camera or something. That is quite similar to that to an area in the digital twin maybe you can get some information about if you used AI product and so on, to get some behavior into digital twin maybe, but it's yeah, it's a challenging question.	DC	PD
00:44:49 SPK_1		
Yeah, I think so, because it's pretty hard to predict humans behavior in that way by using data or.		
00:44:54 SPK_2		
Something. Yeah, yeah, yeah.		
00:44:57 SPK_1		
Absolutely. So I'm interested in the Digital Twin projects. To what extent do you think that is a successful one, that's successful, successful use case of using digital twins on like achieving the goals that you guys have set for it?		
00:45:17 SPK_2		
As I said before, I think it's, in the working process I think it's very very successful to sit around the digital twin different functions and also to understand how the place will be, where you can look at it in a 3D perspective and so on. So I think that is the mainly the two best goals for the digital twin in my point of view.		TCA
00:45:47 SPK_1		
For communicating. Basically.		
00:45:50 SPK_2		

Yeah <u>communicating absolutely. And also you can test the solution that you can communicating between some different functions also.</u> So yeah.		TCA
00:46:02 SPK_1		
So it's like exploring opportunity as well.		
00:46:06 SPK_2		
Exactly.		
00:46:07 SPK_1		
Yeah, yeah, understandable. Okay. So who owns the data that is collected?		
00:46:22 SPK_2		
I think it's a very good question. It's depending on what the data. I think. <u>The most of the data that I've worked with, the traffic data that it's our that that is the that we own that type of data.</u> But in some cases we buy some data. <u>The GPS-data or something, maybe in that case it could be what you say, another agreement about the data maybe.</u> So it's also depending. <u>But in mostly of the cases we are, we are owning our own data.</u> But it's also good question to <Name> about the general data in the digital twin. Yeah.	DSU	
00:47:17 SPK_1		
Yeah, thanks. OK, I have I think, I guess so I want to ask for now. So are there any further limitations or concerns that you see that we didn't talk about yet or didn't touch upon?		
00:47:40 SPK_2		
No, I think we talked about what I what I see and the big the challenge thing is to go from that building a digital twin to work with digital twin. That's the, yeah, the big thing, I think.		
00:48:03 SPK_1		
Okay, sorry, I just suddenly got a question on my mind. So is there any workshop or? How to say knowledge transfer from the technical company and they makes you more understandable about how to digital twins works because you mentioned about one thing, the challenging thing is that how to use the new technology like digital twins and also cooperate with the old system that you guys are using in the traffic department. So is there any knowledge transfer from the technical company to you guys so you are more like used to using the new technologies?		
00:48:40 SPK_2		
In my point of view with all the only thing that we do is we are testing the digital twin in different projects. That is what we are doing right now. We have not set up the networks and so on. Right now we are maybe <Name> can answer that question more but from my point of view regarding traffic flows and so on. The we don't have		

the set up set up with networks, we only tested the digital twin, the traffic flow with the different project just to see and learn about how it how it works.		
00:49:34 SPK_1		
So do you have any question for us, because I feel like we have already asked you most of the question and we got very good interesting answer from you as well.		
00:49:42 SPK_2		
So it was a good, good questions from you, Yeah.		
00:49:46 SPK_1		
Yeah, thank you so much. We were wondering if we come up with something afterwards and realized that we wanted to ask you, is it OK to send an e-mail and ask Further questions??		
00:49:58 SPK_2		
Yeah, absolutely.		
00:50:01 SPK_1		
No problem. Yeah. So I think that's all for today. And thank you so much. And I hope you get well soon. And yeah. Yeah.		
00:50:08 SPK_2		
Sorry for yeah.		
00:50:10 SPK_1		
No, it's totally fine. And we hope that you get recovered soon. Yeah, yeah.		
00:50:15 SPK_2		
Sorry for what I yeah over here.		
00:50:19 SPK_1		
Thank you so much. Yeah, take a good rest.		
00:50:22 SPK_2		
Bye, Yeah. Thank you. Thank you. Bye, bye.		

Appendix 8: Interview 6, P6 & P7

Theme	Colour Codes	Factor	Text Codes
Context	Yellow	/	/
Data Management	Blue	Data Collection	DC
		Data Storage & Use	DSU
		Data Sharing	DS
		Data Quality	DQ
Algorithmic Bias	Red	Data	BD
		People	BP
Principle of Discourse - Democracy Principle	/	Pragmatic Domain	PD
		Ethical Domain	ED
		Moral Domain	MD
		The Theory of Communicative Action	TCA

Participants	
Authors	SPK_1
P6	SPK_2
P7	SPK_3

Transcript	2nd layer	3rd layer
[00:00] SPK_1		
So that we can remember everything.		
[00:02] SPK_3		
Yeah.		
[00:05] SPK_1		
you didn't find a pencil or		
[00:08] SPK_1		
Oh, actually I got a pen		
[00:10] SPK_3		
oh, yeah, we have this big one behind you like. Oh good work.		
[00:19] SPK_1		
Yeah. I'm thinking you could sign the same one. You can do it after also when have the time to read it. Oh,		
[00:27] SPK_3		
yeah, I already. yes.		

[00:37] SPK_1		
OK. Usually we will first introduce also, but we already introduce ourselves in the ground floor. But yeah, but still I'm <Name> and <Name> and we are now studying in Lund University for a master programming information system. It's a one year program. So thank you both of you participating in our interview voluntarily for our topics. So it's investigating algorithmic bias individual twins design for smart cities. So yeah. Yeah, so if you would want to start by telling us a bit about your role right now what you're working with.		
[01:13] SPK_3		
Should you start ? See print name the name of the informant. I'm the informant.		
[01:23] SPK_1		
Yes.		
[01:24] SPK_3		
and what is the print name?		
[01:26] SPK_1		
I guess it's I don't know actually one is		
[01:29] SPK_3		
that I don't fill it out.		
[01:33] SPK_1		
I guess one is the signature one is just the name.		
[01:37] SPK_3		
Yeah, yeah, whatever		
[01:42] SPK_1		
We could use one like for you. yeah..It is just out of formality.		
[01:47] SPK_2		
You.. I will just write on the print name. I write on this side.		
[01:59] SPK_1		
OK. Thank you.		
[01:59] SPK_3		
So now we have a different month. Whatever.		
[02:04] SPK_2		
About our roles here. First of all, I'm not used to say this is English. Yeah, maybe you have to translate. Yeah, it's fine. But I work as a <Role> or <Role> is my role here, but I work quite a work with GIS and maps and our GIS system of a platform is there yesterday? And then what else? We work both we have in the city council you say that. And that's <Public Institution> So we were both within our department but also for the whole municipality personally. What else we have a little bit different roles?		
[03:00] SPK_3		
Yeah, I'm more. So to say this guy [shows programming code]. I'm usually going more in depth in things like okay, we need a program to solve on this or we need to integrate to something and so on there. It's where my part go deeper then <P6> which maybe find out the requirements or what you need and I try to shake out the numbers and things underneath to fulfill and I'm also a bit thrown in the municipality's things about IoT and Innovation a little bit there. So which touch digital twin a lot? But it's not where we put our main effort. So. Actually in our part IoT isn't a part so to say but we have maps and everything where you can apply all the IT data too. ==So we are like more of the base or making with the one that make it possible to visualize in your geographical way.== So that's our part to start with. Yeah, we probably have 100 things. So that's the question.		

[04:30] SPK_1		
Yeah. Yeah.		
[04:33] SPK_2		
We have a lot a different departments to ask like you said, what do you need? And and then we try to we are like the consulate like		
[04:44] SPK_1		
Consultant, capture their requirements and stuff.		
[04:48] SPK_2		
yeah Exactly. And then We got back. And then Maybe we do something special and you can develop it and go back. Okay consult.		
[05:00] SPK_1		
Ok. So would you want to tell us a bit more about the digital twin that you're working with? Is it just one or is it		
[05:07] SPK_3		
it's depends on your definition of a digital twin? Yeah, of course in our definition of digital twin. We've been working with digital twins long before the name existed. because in our view then it could be that you have a data connected to geography that represents the real world. Yeah, and this definition we've been working with digital twins like since before you were born. I don't know where yeah born, but it's all right now but we've been working with this since the last millennium. Maybe not here. But so But if you think more about so it's a little bit what's your definition of a digital twin? You said digital twin? And what was the other part of the thesis thing?		
[05:57] SPK_1		
Okay. Yeah, because from what I understand about digital twins is that we are using real data to kind of like replicating what's happening in the reality and then put it back into physical world and then try to communicate each other. So to provide some decision making or maybe some kind of like prediction and stuff but long Yeah, but long time ago people called it a 3D City. So you just kind of like modeling using Geographic data to model a 3D city out and then try to do simulation in it, but not using real-time data. But it's interesting to see like what's the changes all those years when you are using this kind of things.		
[06:39] SPK_3		
We're we have almost nothing using real-time data and near real-time data. so For our part it real-time data doesn't matter really like so it's light on or off right now. It's not important. It's more like okay. It's been on 20% of the time the last three months that could be anything for us, but it's not like okay now it turned on. Oh, no, it's turned off. That's It's not interesting for us directly. But actually we help <Public Institution> with this kind of things. They're starting to get. This kind of data on the possibility to turn on and off Yeah lights, lighting through API and so on. So we're starting to help them with that. but it's yeah, we haven't. Implemented what exists is that? You can see where all my life and more information on here's a light post that has an ID Etc from this one if we like we can apply. Turns on and turn off. Mm-hmm, but the usual problem here is that we have the possibility to do it. But so far they haven't bought the license for the API at the other end. that's just a playground that we can play with but you can't do it for real because they haven't bought this last part. So I don't say so much real time data. In in our part, there is no we don't need real-time data. We need near real-time data like okay, there's a There now there's a new house or now. There's a new Lot, fastighet (real estate) Yeah. Yeah,	DSU	PD
[08:52] SPK_1		
She will translate it no worries.		
[08:54] SPK_3		

And so on things that maybe happened yesterday or today will we use the next day? The things, but usually we didn't use direct data.	DSU	
[09:07] SPK_1		
Mm-hmm. So you just tried to keep it up to date with the physical world.		
[09:11] SPK_3		
Yeah. And so.		
[09:15] SPK_2		
And so far, it's been good enough. Yeah.		
[09:20] SPK_3		
We're sitting on the <Public Institution> how we build the town. We give permits so to say in different ways to do things and so on so we don't We controlled by Political things so we have mostly input.		
[09:41] SPK_1		
Okay,		
[09:42] SPK_3		
of course. Yeah the output we have it's like sending to <Public Institution> or sending data to other things so on. but we don't have this IoT part. The town is looking on IoT. Like what I know exists. Is like if you go into the bath, there's like is there soap in the soap dispenser or not or we have like four trash cans which can say unfull. so it's what exists is mostly proof of concept. I'm quite sure if you go to the things like water or this kind of things they have more. so they yeah, they are <Company> and power and so on. Of course, they have more IoT and sort of digital Twins. But they are not open. Of course. Of course. Yeah That will not work so good. So from our part. Yeah, we have a lot of data. But we use them one way. So yeah, most digital twin is just one way. It's not settled from that. You use digital twin to directly steer back.	DS DSU	
[11:09] SPK_1		
Yeah. Okay. Yeah, it's not there yet.		
[11:13] SPK_3		
No. Yeah they are exist. Yeah, of course. If you look at it we have we checked in at the door. That's a part of the digital twin. You would like to Yeah, so but it just which level you put the thing is, so or part of the whole thing. Now you have to ask more questions.		
[11:38] SPK_1		
Yes. How is it used like you mentioned permits and like what do you do with the digital version of the city. like the goals of like you want to achieve by using your definition of digital twins.		
[11:55] SPK_3		
It won't…ask the question once		
[11:57] SPK_1		
Like what's the goals? I'm thinking first. Just what are you doing with it? Like because I'm not really understanding.		
[12:03] SPK_3		
Okay. Number one is decision making yeah, should I show them?		PD
[12:11] SPK_2		
Yeah,		
[12:11] SPK_3		
for example, should we take <Project Information> because that will be more easy to understand from normal people. We have a lot of maps with different what's here?		

[12:33] SPK_2		
so the front departments or anywhere with the special of seeing them, you know nor your project you have your own.		
[12:42] SPK_3		
So this they are using for if you show that permit to build something so, okay. Let's go to here. So so then I can jump directly to this building or this lot here. So so here we are, <Project Information> hasn't update the this is our we were on the way to migrate to this map. it looks exactly almost exactly the same as the last one but So if you're here you can. We use it as a hub this one thing. So here there's something called the <Project Information> in Swedish. This is what's decided. How you will be able to apply for building something. So if you're just click here I can go directly and then I get the document. which tells what to do and so on. So that's one thing. Then we have more things about this property so So if you would like to know which addresses exist here. Who owns it get more different documents which is connected to this. Yeah. You can get a lot of this kind of things here. What you think or what I think is quite funny. Is this what we call <Project Information> This is a we took a lot of All our data and then we made a big script. that looked at it from many different perspectives and made a lot of micro analysis on it and then I built a GUI to show it. So for example, we can because when you if you should be allowed to build something there are parameters You look at for example, if you're closer than 30 meters to the railway. Yes, we are. and here is some information that we are in a What's defined as kulturhistorisk värdefull miljö So yeah, actually before this was built here was things. that were We should save there's not a right here today. It's not a relevant, but And there could be parts that should be cultural still there	DSU	PD
[15:38] SPK_2		
could be the building or the land formation.		
[15:42] SPK_3		
Yeah and so on.		
[15:42] SPK_1		
Okay		
[15:43]		
so and here yeah, this is usually not standskydd. It's a beach protection. So there's not so you don't have any water so we don't have to think about that because it could be like you're you're allowed.		
[16:01]		
In one regulation to build the house here, you can get the right to build a house, but due to strandskydd. You're not allowed to build it. so it's good to know that we can give a permit to build a house, but you can't build it because there's a different law that say you can't build it there. And yeah, so there's this kind of things and here we just cherry picked out from data like auto photo from different year's. This is just more how the property now the boundaries is, this is not interesting. This is what <Project Information> it's yeah. It's a different document. So to say which is more abstract tell about it, but it's not so interesting in this case. Here's more. What they measured So this sort to say measured data And this is yeah, I could kulturmiljö. So here's the blue part. It's actually a culture things from and here we have a lot of other things. Yeah, so here's The document you saw before with the information. Here's on the cases connected to this one and you can from here you can jump further and see like, oh, I would like to see the bygglov for yeah	DSU	
[17:37] SPK_1		
so this is like building permit.		
[17:41] SPK_3		

one this should put up the then you sign here. So here's a paper for that. So we use it a lot like a hub and integrated with a lot of other things. I also integrated with <Company> so you can say actually directly on the right place on <Company> so you can Yeah, see. So here's one thing where we have on the data. We put it together. To help them take a decisions	DSU	PD
[18:13] SPK_1		
About building permits. For example.		
[18:16] SPK_3		
Yeah, <Project Information>. Yeah building permits and then we have for many other things. We have for <Public Institution> [inaudible], probably something [inaudible] in English to Then all cut off or cut off they ask you something.		
[18:36] SPK_2		
To Survey?		
[18:42] SPK_3		
Yeah, so they yeah. And we have for school for example for school there I build. It's not only maps I have. Okay. There we have real time analysis. If you apply for a bus card somewhere that you fill in And then when you filling out, I want to have a bus card with it say oh you have only had a <Number> meters to your school. You will not get a bus card. we have like you calling and make in real time analysis. Sometimes and also for which school children should be allowed to go to. So or not allowed to go to which order you should get prioritized into the school. If too many people would like to go in the same school and there's things that has to do with the distance and sounds so we have to calculate that. So this is the kind of things and or back to the question about the digital twins. The next step we're going now.	DSU	PD
[19:50] SPK_2		
Because this is more like so the paperwork with the this is instead of oh they forgot about just checking Is there a beach protection? So this is like standardized and making analysis instead of doing it more.	DSU	PD
[20:10] SPK_1		
Manually.		
[20:11] SPK_2		
Yeah. Yeah, so it's more like making the decisions safer you say more time efficient. So that if we work at the same thing. Yeah, but just in different areas the same way. it's not up to	DSU DS	MD PD
[20:32] SPK_1		
To standardized the decision.		
[20:35] SPK_2		
To more standardize.	BP BD DS DSU	MD PD
[20:36] SPK_3		
Let's see or maybe how… what do you have it here?		
[20:41] SPK_1		
Oh, there's traffic information as well. Is it traffic?		
[20:48] SPK_3		
Now I ..this here was a cultural building which so if you if I point out building here. Then I can there's a lot of information here. Let's see here. Yeah. Perhaps this works. Yeah, so then you coming to a different source, which has information about that	BP BD DS	PD

building there? So it's we usually most people use as a hub this way. To get all the information we can get a lot more information and we have Yeah our own analysis that we built and so on. So what I know here this for it's not so interesting inside but it's like what is. So if I do turn this on and off. well see, I should turn on and off a lot of things here. Otherwise you will not. so so this is their existed different rule if there's many houses close to each other then you have to. If you should build something you have to take think about that. Yeah, and before it was a little bit like you asked someone how I will build here and look at the map and Ah, it's too many houses or it's too close. Yeah, so you have to apply due to this rules. Not this Rule and then we made an analyze of our data and build this layer. So now now they can directly say that okay, we will apply this rule or not. So otherwise they could maybe call three different people and get yeah, two three different answers. And so this is just a like an analyze on building distance addresses. So we like so we do like this kind of things too to help to pre-build decision-making.	DSU	
[23:03] SPK_1		
So all the requirements for building permits are in there already.		
[23:10] SPK_3		
Not all but a lot of them.		
[23:13] SPK_3		
So more than I think more than many other municipal has. I think that we have more in our maps and if you go to <Location Information> or maybe <Location Information> to I so		
[23:28] SPK_1		
okay. Yeah.		
[23:31] SPK_2		
But it's also way like we have nothing , something that's called <Public Institution>, on the first floor. Mm-hmm, like the first line support and citizens and yeah. thanks to these maps that can answer the quick questions and doesn't need the specialist to answer. so that's one thing as well. So we have a lot of different so it's according to what what kind of information you need.	DSU DS	PD
[24:04] SPK_1		
So it's also used for communication with the citizen.		
[24:05] SPK_2		
In citizens they also have maps on <Website> to show it different..		
[24:11] SPK_3		
They're very specialized a little bit. Where can I walk my dog or I thought of I'm elderly I would like to go out eat with elderly people where is the…so they're very they're very focused.		
[24:28] SPK_3		
And here there these These are focused for task but the task is quite big.		
[24:39] SPK_2		
but it's also like where they are <Project Information>. And that's then they use these maps to see where is it possible to move where do you have the land. that we can move. Is it big enough where if you <Project Information> and where there's other activities going on where they're supposed to move them. So where do we put why do we know the other things and who's today using the <Location Information>. Is it like the tennis. where does the people live that's go to the tennis courts today and yeah, okay to for the future. Okay, and then they can test like it if we move here then we can see you. you're supposed to have like <Number>		

minutes to the commuting. so it's more like instead of building and then see if it's okay, but you see here.		
[25:46] SPK_1		
so it's like simulation things		
[25:47] SPK_2		
Yes		
[25:47] SPK_3		
Yeah. but the quite typical things this will how far from a bus station are you like I've made analyzes to tell where should you live? If you should be closer to than <Number> meters walk the way not bird to reach bus station for example and so and then we also made we took bus data the And how the classified things say, how many times per hour the frequency of what's a frequency this bus stops here. So the frequency and the distance and then classified into different classes and then you can like then you can say is it the is it good place to live for using public transport or not? So that's kind of things we have done and but I don't think we have it in any map here or I can remember.	DC DSU	MD
[26:53] SPK_1		
So if it's not a good place to live for using the kollektivtrafik the public transport, do you then try to fix that somehow or what's the…	DSU	MD
[27:04] SPK_2		
Yes, there's our		
[27:06] SPK_3		
this was for the plan or it's more like where should the bus stop it goes like this or should he take a different road? and we're also did it for elderly people. <Project Information> ? But yeah		
[27:33] SPK_2		
but if you know, there's a problem then it's easier to to build the future just minimize them.		
[27:42] SPK_3		
you may be confined strange barriers succeeds looks so if you take the old classical bird way, oh, it's very very good. Then you okay learned how's the railway? It if you look here in the middle, you have a highway so there exists. If you live here on the bus goes from here. It's only 80 meters away, but it's four kilometers to go there, so therefore it's good to make different kind of analysis to understand and so on.	DSU	
[28:20] SPK_1		
Who is included in this project? Who's participating stakeholders?		
[28:30] SPK_2		
And we which program project is?		
[28:34] SPK_1		
Okay, so it's different projects or I mean in these kind of planning decision-making.		
[28:40] SPK_3		
And the planning of the decision making it's almost the whole the municipal. but in in the map part and visualization of the maps and so on we're <Number> people soon for if you're lucky that is like making it possible.		
[29:01] SPK_1		
People as like individuals. Okay. Yeah within the municipality.		
[29:07] SPK_3		
Yeah, which is so little bit too little. Yeah. and we're making like everything from that the program is working to what you see and so on so a lot of things		
[29:23] SPK_1		

I was thinking like businesses other companies that they are not included somehow		
[29:31] SPK_3		
They are indirectly they're using our public maps to get the information and we're yeah if they're building in the <Project Information> or something Like and then they we are having other people coming in in a help us of course and in some way. They use the data what what we actually did we made. It's not in this map. We made things so you can order data automatically without receiving it more or less. So do I have anything like that? Okay, this is I'm just trying I'm just trying to make it a newer thing of this now, but we have like I'm interesting in this part here. So and then I can say I like to have a I need something for AutoCAD here and then I can order it and then I get an email then they down on the result. if you're going to other not a simple answer. It's a little bit like they call it. Yeah, I need this data here. Can you make an export of this? So this thing makes that we can be two persons less. Yeah, because then you can just yeah instead of we should try to understand which square do you need? Which data at which format? We precocked things in the maps so they can get it themselves.	DS	PD
[31:07] SPK_1		
Okay? Okay, so they consume the data, but are they also producing data somehow for you or it's every day like all the data is it you that collects it somehow		
[31:17] SPK_2		
You, as the municipality?		
[31:19] SPK_1		
yes, I guess or this. Or from others and department. Yeah institution like traffic department. Are they also providing those data traffic department or something like that?		
[31:31] SPK_2		
Yes they are.		
Yeah. Okay. So it's the difference department. So it's not yes like <Public Institution> and also we try to get information like if you build a school and maybe <Company> I don't know and they will try to get The models from and..	DC	
[31:51] SPK_3		
that's what we're trying now. That's our next step. Yeah.		
[31:57] SPK_1		
so they're the <Project Information>		
[32:00] SPK_3		
now, that's that's the process is you have <Project Information>, which is very abstract, right? Yeah, and then you make the <Project Information> which is less abstract. then it's coming down to the one that really would like to build something. Yeah, and then they look at the higher document what it's allowed or not. Okay, so to say if you make it easy, but We <Number> don't produce any. How to say base data. or but what would we do we produce analyzes from this if you take this data, so last analyzer did I took tax information and so on and I build a Layer, which could tell where do we have houses where many people live And in which of these houses? Is it the bostadsrättsförening from not possible to translate because I don't think it's there and which are not and then I can visualize that but this is like then we collect data from <Public Institution> which actually is we have a kommunal, it's technical people by us that's produced the data and then we take from the <Public Institution> information and Yeah, I think it's these two things that we use is here formally. Yeah,	DC DSU DS	
[33:31] SPK_1		

so not that much private entities. Or private companies more governmental.		
[33:37] SPK_3		
No, we can say when we get data from other parts, then we use the services. So if you take a skill or something like that, if you know what <Name> is	DC	
[33:51] SPK_1		
Is it a company?		
[33:53] SPK_3		
No no it's		
[33:54] SPK_2		
Swedish geology		
[33:56] SPK_3		
So if you're interested in what kind of mud or dirt or stone. Yeah instead of trying like we're using services to show this kind of data in our maps so we have a lot of data which like we're streaming from others. So when they look like we use it from the source if we can and if we can't we're downloaded it from the source from for our area and put it on a server. So there it's fast to get it but from there's I don't think I can't see anything that we buy from private companies such thing Yeah in some extreme. indirectly some small part, but not not really in this kind of things now not today.	DC	
[35:01] SPK_2		
But we do sell it or we have it that open data, some of our data that are they used like	DS	
[35:08] SPK_3		
We're we're here trying to get more and more open data. And in theory, the municipals try to get open data to you, but it's not like the politician has understand what it means and how it used and how it can. In the big term be your own money by having public data. We definitely mean at our part, but the problem is that if we have open data. We will get a higher cost to have it and other people will earn from it. So it's it's all yeah,	DS	PD MD
[35:45] SPK_2		
but it's changing now.		
[35:47] SPK_3		
Yeah, it's changing but usually it's like we have to produce today. We are our models is that we get money for produced data. Okay, and then we can't really have it. Or we sell it in the other end. So that's the way we have money to produce a data, but if we can't sell it. come with them produce it then you taxpayer have to pay us to produce a data but in the other end. It will get more work done and get more tax money back. So to say so it's complicated question.	DS DSU	PD
[36:25] SPK_1		
it's so for the data is most likely about the geospatial data. There is no like human or sensitive data involved it in the models or the twins, I would say		
[36:39] SPK_2		
What do you mean human can you specify like		
[36:43] SPK_1		
like people movement or like traffic flow is maybe yeah.		
[36:49] SPK_3		
Yeah, we have we have some kind of traffic flow but they're not live. It's a little bit it's students that's stands in a corner like one two, three …five. and then we aggregate that for a whole year and then we put it in.	DSU	
[37:09] SPK_1		

Okay, so it's not identifiable or sensitive data		
[37:12] SPK_3		
No no.		
[37:13] SPK_3		
but okay we have We're having data connected to the tax and the owners of properties. So yeah, we have Information that is yeah GDPR and So on, and of course we have when you apply for something. We have your many of your if you apply for permit to build then of course, we have your name and the things and we have it connected to map but it's not. Easy, it's not. If not allowed to see it, you can't see it. so to say.	DSU	ED PD
[38:02] SPK_1		
I'm thinking about you said that you have students that counts counting people. Yeah, it's like is that how do you decide where that's needed? And what do do you make decisions based on that kind of data?		
[38:18] SPK_2		
It's that <Department> that use it for planning or		
[38:23] SPK_3		
but what could be interesting for your things here is that we were talking about light		
[38:30] SPK_3		
poles.		
[38:31] SPK_3		
Now, <Project Information> which means that you can put on sensors on light poles. So for example that in a near future we could have a sensor that counts cars bikes pedestrians.		
[39:03] SPK_3		
Maybe how fast they go which direction in real time data so we can get it in and then but as I say for us we don't care about real-time data and our position usually but if we aggregated and say, oh here's a lot of people like okay this way in the morning of all it's using in one way, but they don't use in the other way. Where is the people going? Why is it or we can find out patterns maybe in the future, but and the real direct data, we don't have so much interest others may have like so I think if it looks down on the road, there's I think there's something that that counts bikes that going by. but but it's the we're just so you can count bikes going by	DC DSU	PD
[40:02] SPK_2		
on the way up if you look at.		
[40:04] SPK_3		
Unfortunately, both in from outside so you can look at the data, but it we don't have it integrated to our whole system support. But what we aim for is to get more and more of this data integrated so so we can have them in our maps more more but I will say the more integrated and IOT things will get in the less public today. The maps will be.	DS DSU	
[40:44] SPK_1		
So there's a trade-off.		
[40:45] SPK_3		
Yeah, but the municipal at … the big is looking at IoT how to build the infrastructure for it and there they more looking at this how to make it more safe and how to you can get IOT data and many different formats how to harmonize it for use and how to store it in silos in the middle and sound. It's they're looking at that part And then we're like a higher part of it or maybe a deeper depending on where they go, but	DC DSU DS DQ	ED PD

they've been doing this for at least \<Number\> years, but there's no real thing right now.		
[41:39] SPK_2		
It's again more like the proof of concept. like		
[41:45] SPK_1		
so when … For example like the case that you mentioned about like planning for maybe \<Project Information\>. So when you're doing analyze analysis and doing a lot of simulation, maybe it comes out a lot of good option then how to weight those option. And what is the process or discussion involving in there to reach out the consensus like which option to pick?		
[42:11] SPK_2		
that's the specialists		
[42:18] SPK_3		
It's it's actually usually not up to us.		
[42:21] SPK_2		
yeah it's just tools and then they have like their knowledge, we have these to choose but according to what we have been before and we know that this will end up in the best solution.	BP	MD
[42:35] SPK_3		
Okay, or they say that oh, we we doing like this then it will be like so we doing like this and then it will be like that then we do a simulation and say no that wasn't really true then it's up to them if they will. Yeah use it or not. It's not our thing to decide if quite okay. It's annoying if you calculated and see that what you're the decision you're making is based on wrong data, but it's not up to us. to say how they take decision And it's I think in a lot of part where we working it's very hard to quantify things with the unit and say what is good and what is bad so. I can take a example where the buses I'm living in the next municipal and I realized that they removed almost all the stops for a bus to \<Location Information\>, which will make me one and a half minute longer walk to get to the bus, but now the bus to \<Location Information\> is faster than the bus to \<Location Information\> even if it's longer distance, So In total, is it better? Allows her for me that so they're analyze we made before is a little bit. Oh, it's bad because it's longer but it's maybe easy to park the car or so. So what actually happened in one analyze. It looks like it's not larger better, But in many use cases it's a better solution. So it's you can't just look at that parameters that you get out.	BP DS DSU	MD
[44:30] SPK_1		
Mm-hmm. So it's the experts that makes this decision or yeah, so you're		
[44:39] SPK_3		
Not we're not making any decisions. But we sometimes sending Providing quite clear data that	DQ	ED
[44:46] SPK_1		
to supporting their decision making yeah,		PD
[44:47] SPK_3		
Yeah, exactly		
[44:49] SPK_2		
Sometimes we make the decision or we show it to the politician. Yeah, and they can make the decision, right? Okay, because it's it's the way it Works.		
[45:02] SPK_3		
We can take the example with the the bus card that that's a quite easy example to understand. if you fill it out. So it actually says that due to the rules we have you will	BD DQ	ED MD

not get a bus card, but if you like you can still send in the application to us and say I like it. Anyhow, because yeah, they are they are whatever so but in most of the cases they will of course decide that you don't get it. So it's not up to me to decide if they will get it or not. I just make a pre assumption that due to the rules we have you should not get it. But yeah, it could be that the calculation is wrong. And now we have had a yeah, it's that one there. They were <Project Information> or so in <Location Information>, which made that many people got <Number> longer distance to the school. And how should this affect so when I actually start in the school, the tunnel will probably be there, but when we calculating the distance because we calculating it on the data today. They will have a longer distance to the school what will happen So Yeah. it's it's Yeah, you can't blindly trust data, but it tells the truth. From the input of the day what you put in then you get sort of the truth back and then you have to understand that. Yeah, how are you thinking about taking the decision in The future? That's yeah,		
[46:57] SPK_2		
And that's important to tell the one that's asking for the analysis to tell them. That's the data but we know that this this so maybe still not a good analyze. Yeah. Okay. So analyze made the way they said but it maybe doesn't fulfill their purpose so	BD DQ	MD ED
[47:20] SPK_3		
So you also give them some kind of indication of how like solid there.		
[47:26] SPK_1		
Yeah. Yeah. Okay. So is there okay to show?		
[47:31] SPK_2		
Also have to show it on the map like the result is not like the map it's the final thing but it's still like it's important to make it clear. So the one that gets the information is easy to understand. Yeah	DQ	MD ED
[47:46] SPK_3		
From the digital twin and the city part what we're starting a project in is to more visualize it in 3D as much as possible and also try to collect how to say future possibilities how things could be in the future? Because if if someone asks for say and <Project Information>, then they say I need document that make it possible. To build a house and make it possible to build this kind of house. So then we can think like yeah, if you take bigger things like if you look at <Location Information>, so they may be it's if they when they applied for the document that they should change the way this piece of land should be used. Then they can say here will probably be a very tall building. So then we can visualize that as a future scenario on our maps. And after that someone asked for the permit to build it then it's like okay, it's quite certain that this will happen, but we can like pre-visualize it that was you said that we can get in models. So this is what we're trying to do now. Before all this model was a PDF or a paper initial whatever And now we're on the way to trying to get these in. To get a visualize visual way of looking at things And into the future and different scenarios. So that's the way we're going now. I hope it will be good. and how we have it today. It's like we don't have the building. We have a dot. So to say today we're represented with a dot in two different colors. Say more or less or we can do it more like that. And what we will try to do is to do it more in 3D visualize and so of course then you also can do sight. So if I stand here will I see the church or like that or will the will this building throw shadows over my house? So		MD PD
[50:26] SPK_1		
Yeah, I was thinking we have some different. Okay, if you .. I've I got a question but it's kind of like slip away of it. I go with your question. Okay. No, because when we're talking about data like building the 3D models and stuff. So how do you feel		

about the data quality? Is there any like discussion like you talk about? How could the data is? How is it presenting the real world? Like you said the building is now as a dot how much a level details do we need to represent in the 3D models? Is there any discussion about...		
[51:00] SPK_3		
So did you have 2 hours or did you have a week		
[51:03] SPK_1		
is there any industries standard that is following for like collecting data. So using the data.		
[51:11] SPK_2		
There will be.	DC	ED
[51:13] SPK_3		
I think the problem is if you have too high detailed future scenarios and people think that this will this is the truth. So therefore we can't have too high level of detail things that doesn't exist. And so and yeah, it's so there you really have two more or less try to make it sketchy or something because otherwise people will think it will be like that. and also if like, oh, we made it quite nice and then it's slightly wrong color. And then No, no. No, this is not what what we're going to build. people should not think like if the company sent in a model and we change the color from a different. So just a little bit and then they can complain why so. There's a lot of things but To I think we have to try to keep it the level down If we go to <Project Information>	DQ	MD
[52:34] SPK_2		
But it's true.		
[52:34] SPK_3		
I did some things that take data put it in to <Platform> and yeah, we can do it and may		
[52:50] SPK_2		
Just because we can do it we won't do it. It's like you what do you we ask what do you really need the <Department>. What's the need for you. It's not like I can do this and this it's more like from the needs we build the 3D model, the digital twin. Did you and so it's it's not that visualization <Project Information>, we don't have their resources not anything to do that. So it's more like Yeah in another it's more data more information.	DQ DSU DC	PD
[53:29] SPK_3		
In one way that they can. It can be good and sound. For example, I think they did <Project Information>. It was the way it looked before so on, okay. You can do it like that too. But that's not the way that we aim. We may take a little part like okay, we should ask people and in the schools for something and then okay, we have the data. Okay, put it in <Platform> so they can have this around and yes here would like to have something here would like to have you here. I see when we they ask for inputs from younger people it's been used in used in other parts. So perhaps from our data that we fast can generate this for a use case maybe But that's not our goal. It's it's fun. It's like it's the selling point to look at the YouTube videos or so on that's like oh this is the way but we're not aiming at that one. you will pay too. You're not interesting to pay the taxes for we should do that.	DSU	PD
[54:57] SPK_1		
But so you're trying to incorporate feedback somehow from citizens or is that a future?		
[55:03] SPK_2		

Yes, we do that already, actually. Yeah.		
[55:06] SPK_1		
How do you do that?		
[55:08] SPK_2		
It's a democracy and We all the different parts it's like the <Project Information> this also to the citizen to get their feedback. Of course they did it together and then you put it together and according to these opinion. We need to change this. Okay,		TCA
[55:26] SPK_3		
There's many digital Integrations between people don't think so much about the digital twin and GIS part for that. But For example, we have integrated things in our system which can sort of publish things in where things should be published in paper. No one reads, and I forgot the name of it but and we also have like all the municipal say someone should build something at one part. We have the information about every body around and we can send the mail to them. Or we can even send a <Project Information> so you can send digital or analog information to these people which living close and have something to say so that and I don't know if when you went by there's public papers of maps. which you can say, oh I'm in I don't like this park here. That she built so usually there's <Project Information>. I don't know if that's someone right now that you can say anything about	DS	TCA
[56:48] SPK_1		
on the website or <Project Information>		
[56:51] SPK_3		
<Project Information>. there's papers in the windows and you can find them on internet too and we have		TCA
[56:58] SPK_2		
We have to post them in different forums. Yeah.		TCA
[57:03] SPK_1		
Test for feedback		
[57:04] SPK_2		
Yes for feedback during special time, like this is for now. You have like a month to give us feedback for <Project Information>		TCA
[57:14] SPK_3		
We have the digital services that so you can go into one service and if there is anything that you can say anything about you will get here's <Project Information> . You can say anything about and then you click and then you can send in information or send in your complaints good or bad. So there will have integration with GIS and our systems and our eServices. This is there actually.		TCA
[57:50] SPK_1		
Nice. That's interesting I see that we're a bit over time I don't know or Do you have something to.		
[57:58] SPK_3		
oh, I don't have anything. This is interesting.		
[58:01] SPK_1		
Yeah, Cause we have I think we have some more questions, because I want to ask this. Yeah exactly. We want to ask what's your current understanding of algorithmic bias?		
[58:21] SPK_3		
So what what's your definition of algorithmic bias?		
[58:23] SPK_1		
do you have a thought of it		

[58:27] SPK_3		
I think I have a thought of it but..		
[58:31] SPK_1		
If you want to like... cause we're interested in how like what you know about it and what's your understanding of it? So if you want to try to explain if that's not too much putting you on the spot.		
[58:46] SPK_3		
You said I algorithm		
[58:48] SPK_1		
Algorithmic bias.		
[58:51] SPK_3		
I can't. Algorithm. Let's see here [search in google] I think yeah. Okay. Yeah, shit in shit out. No, I'm really aware of it. Yeah. And but okay, this is a tricky question because In some way. I think this stops us Because we have the too much knowledge about it. Which say that oh, like I say, it's our in data is 95% correct? So no, we don't do it. We don't do the analyze. because we just it stops us I think the problem is that it more often stops us to do analyze and if you look at the If you have a data source. And it costs I'll take 100,000 to make 95% correct? And then it will cost a million to have a hundred percent correct, right? Okay last five percent the last five percent to make correction is always the most expensive. And in our way, many times it's a little bit like if it's one thing wrong, it's a bit like oh, we don't do the analyze and but I think we should do the analyze and how the understanding of there is errors in the data which may which makes usually errors in the output but you will get a quite you'll get a bigger picture of the whole thing. Even if you have a bias in the input, so but sometimes it stops us and many times what actually happens in real life. It's like you have a very very good analyse and there's a point which is bad and then I say they throw everything away. Of course, they can say they can say, oh, here's the one error then everything is bad. So yeah, there's a it's a two-sided coin. And I'm a <Personal Information> in the beginning. so it's for me it's hard to produce. Yeah, I know that this isn't 100% true or it's it's true from what you put in and you get out and but I know that it's wrong but in the other way in many ways it can give you a better thing to take decisions from	BD BP DSU DQ	MD ED PD
[01:01:43] SPK_2		
Because it's understanding		
[01:01:44] SPK_1		
One you still know.		
[01:01:46] SPK_3		
or you can get a harmonized way you take decision from the thing. I showed you before with the map of sammanhållen bebyggelse which like here. We apply a different rule before it was like you look at it and a person should every time make a decision and look at it. Now. We made a computer do an analyze of it. And Every Everyone I showed it for in different municipals. I tried to start using the same thing. And yeah, this is better than the thing we made by hand before. And it's also it's same for everyone so so in one way it's, yeah if you look at this that before it was people that took the decision. And you you can get free three decisions. Now you will at least get the same decision. So yeah.	DSU DQ BP BD	MD PD
[01:02:54] SPK_2		
We are aware of the problems cause it's equal overall.	BP BD DQ DSU	MD

[01:02:56] SPK_3		
Yeah, we're aware of the problem. What we also see is that other people isn't that aware of the problem as we of course people think that okay. Here's if there's a line would say here's the border for his two properties people think yeah, that's The map is right. Yeah.	BP BD DQ DSU	MD
[01:03:22] SPK_2		
If it have it done map sometimes it's like it's on a map then it's the truth.	BP	MD
[01:03:29] SPK_3		
so first of all like all this can be very historical data. It's been transformed by bad mathematical way whole thing. Yeah, it can be corrupted in the way a little bit and so on. and then you maybe have this auto photo like you have a picture from above but if if you have an airplane here and you take like 45 degrees down it maybe looks like this house is an different place on the map. And then you have a line. Which say? Oh look you build this on my property. No, this is due to yeah, so I think we have we have a lot of knowledge about it and we have to yeah, you can't have 100% correct data. Of course, then you will never get any you will not be able to build anything but yeah, a lot of data is quite high and we also have data on quality which helps?	DQ BP BD	MD
[01:04:38] SPK_1		
quality on the of the data then or quality of		
[01:04:43] SPK_2		
Both Both like down to the point and line. but also like the data set.		
[01:04:52] SPK_1		
Okay, so it's it's some kind of metadata. Yeah stuff like that. Okay. Yeah		
[01:04:57] SPK_3		
two levels if you have like this data is measured two years ago. So and we can say that this point is within 10 millimeters from it's real position and or this point for the houses is accurate. But actually today it could be that the house is someone has tear down the house but it's still on the map. Yeah. So so you have many things to know about you how you know when it's collected and how exact of course it's a little bit we have two data sets. This is about buildings which are trying to get better data for because we have high measured quality data on buildings and some buildings and then we have newer low quality measure But what actually happened is that half of this house was teared down. So it doesn't matter that the quality of the data is high because it's The house isn't there anymore, yeah, so it's a little bit Yeah, there's yeah. We we know a lot of things about the data and you have to take that in aspect but in many times I think we take it too high aspect? That was depending	DQ BD	ED
[01:06:25] SPK_2		
Depending on what our background. Yeah.		
[01:06:29] SPK_1		
And Technical yeah.		
[01:06:33] SPK_2		
I used to work at <Public Institution> and then It's really really important. Yeah, the qualities everything.. But also it needs to be updated that's for us. If you're talking about the digital twin or is more important that it's updated that that's when you measure data if it's like one centimeter wrong, it really doesn't matter but is it still there? Okay, I know that. next week. Yeah, that's that's more important. So that's we're trying to do more of that of up still because of different decisions. We need to have the accuracy, for different decisions. So it's important to keep that in different products. But in some it's okay	DQ	MD ED

[01:07:26] SPK_3		
Nowadays. We're before you were usually measuring things in millimeters things. Now, they have more abstract things like well-being. So so if you say I need millimeter data to calculate something that has to do with well-being and is it actually do we need it or and so on a little bit you have to yeah. Say we're measuring more things nowadays, which isn't on and off or so on we try to find other things to and then it's maybe not that important all the way. But of course we have many things where people don't know about the quality of the data. I know one data set. We started collecting it five six years ago and then it has some values for the back and then they took this data to make decisions. What will happen is that it looks like oh, it's go. It's going from a low level up up. Of course. We have a full data set from one year and the older data is like going down line here down four or five years back in time, which would say that. Okay, we have an up the trend going up when we building up the data and then it going as a so because they didn't. We collected the data for one thing. They used it for another thing <Project Information> and like okay, but you don't understand how the data was collected. So therefore you made about decision on it or may take a bad decision. You have to and that. Yeah, it's the data is actually not wrong. It's correct, but it's not the full data set, you know earlier stage and if you aggregated, so, yeah, we have a lot of Yeah. It's hard what to do.	DC DQ DS DSU	ED PD
[01:09:46] SPK_1		
So you're trying to explain that for the decision makers as well. And the quality of the data set.		
[01:09:54] SPK_2		
Yeah, and the one that's making the analysis, of course, we are making but there's different people out on the different departments in the analysis as well. So it's like if you want to use this and this yes be aware of this	DQ DSU DS	ED
[01:10:09] SPK_1		
okay, and I was thinking about you said that if it's I don't know if that's a great number the 95% accuracy or not. Now if that's what you're always. Yeah,		
[01:10:21] SPK_3		
but I know that the last the last part is very expensive to yeah, all the time	DQ DSU DQ	PD
[01:10:26] SPK_1		
but I was thinking then where you draw the limit for which analysis you use based on the quality then and how accurate it can be. What like, what's the decision making around that like, we're who decides where to draw the line and why do they draw the line? What's what's the like challenge if that's not. it doesn't. Do I make sense. I don't know.		
[01:10:52] SPK_3		
This is okay. Usually I think we draw the line and harder. But In many cases, we may say this. This is a good analyze and then someone find a spot and say we don't use it. which is Little bit stupid like okay, there's one error in the data set so I can prove that it's not right. So then we don't use it. So it has both sides and	DQ	ED MD
[01:11:24] SPK_1		
but what's the reason then to not use it when one point is like what's the reasoning behind		
[01:11:31] SPK_3		
people are scared?		PD
[01:11:32] SPK_2		
Yeah I think so too.		

[01:11:33] SPK_1		
okay because they find error?		
[01:11:37] SPK_3		
Yeah or		
[01:11:39] SPK_1		
what's this what what are they scared of do you know or do you have an idea		
[01:11:43] SPK_2		
To making decisions That's not good	ED	PD
[01:11:45] SPK_3		
or that someone can point. Yeah, it's hard to point out. Look you still you saw it there you did wrong. and Yeah, what. so but or the data set.	DQ	
[01:11:59] SPK_1		
but is it an issue of negative outcomes then from the decision making or what? Why? is it just the fact that it's not reliable then or one, you know, yeah,		
[01:12:12] SPK_2		
I think more like that.		
[01:12:14] SPK_1		
So they are more reliant on the accuracy level. They want 100% accuracy to make decisions.	DQ	PD
[01:12:25] SPK_3		
I think the problem is...		
[01:12:23] SPK_2		
It's very different people to I'll say		
[01:12:28] SPK_3		
I think the problem is if you have something on paper pictures and so on. It's not a problem, But when you have it so exact. How it comes out in our analysis. It's a little bit scary. and there's many things you have. Yeah. It's hard.	DQ	ED
[01:12:51] SPK_1		
Do you think there's a like a problem with trust in the technology or? Do you think people trusts that technology that the decisions are made from?		
[01:13:08] SPK_3		
I don't think that they think that there's technology involved. it's so it's so normal now, I don't think that when I open the door that I'm not using the key. I'm using the blip. It's like So that I don't think but in the other way it's a little bit. If it's on paper in the computer, they usually think it's correct. They assume that this is the data And I think many don't dig down and try to understand as I said with this, okay have a data set. Oh, I make a report out of it. So yeah, but you didn't dig down you you don't know that the data set is just correct for the last six years and then it's every data in is correct to but it's not all the data for earlier years and People. Yeah. In general we have a problem with metadata Because we're not actually owning the data and the one that producing the data know about the data. So why should they write metadata? And..	DQ DSU	ED MD
[01:14:23] SPK_2		
If they are the user the end user, but if someone else is using their data like like of course, they should use [inaudible] data. Yeah, then they don't know the quality of	DQ	PD ED
[01:14:34] SPK_3		
and they have hard to share data. They don't think that others can use this data for something else. And so on yeah, it's and we don't We don't have a or almost no one	DS DQ	ED

Text	Col2	Col3
in Sweden that we talked to in the geography world has a good way for producing metadata and to be used.	DSU	
[01:15:01] SPK_2		
Yeah, and that's a lot of standards. Yeah, like the way should write the meta data down. Yeah. It's like so many different, and there's there's a lot of columns that you need to to fill in and there's a lot of them. Yeah because it's too much and we will really need to fill all this information and	DQ DSU	ED
[01:15:31] SPK_3		
even if you're fill it in. Where where will they find it? and how should they come? Look when you look at the map there? There's 400 sources. So, how should you say? Oh this line here. Comes from that source and here I find the metadata. Oh, it was built on a different data set which has this major difference there. Okay, so it's Yeah	DSU DQ	PD
[01:16:00] SPK_2		
and disadvant today on the whole set. Yeah, or is it like the different objects?	DSU DQ	PD
[01:16:05] SPK_1		
Yeah, the linkage between those data. understand		
[01:16:12] SPK_2		
so I don't think we couldn't really answer your question. Okay. It's a trick one. Yeah. Well, it's Yeah. And I think we will like to commute. We would like that other people understand it more but	DQ DSU DS DC	ED MD
[01:16:34] SPK_1		
The bias part. You mean		
[01:16:35] SPK_3		
pardon		
[01:16:36] SPK_1		
the bias part you mean		
[01:16:37] SPK_3		
Yeah. Yeah, but it's It's hard to communicate and sometimes it's it just blocks yourself because for the analyze you should do it doesn't really matter. so like yeah,	BD	
[01:16:55] SPK_1		
Do you think there's and over emphasis on accuracy and too little on emphasis as you said how the data is collected and like do you think if the accuracy was high enough, even though the data is not collected in an even way say like not Accurately. Could you still use that for decision making then or do you understand my question?		
[01:17:24] SPK_3		
We can use the data but it's much more work. When when the quality is changing along? I have a Yeah. Stupid things that they asked us for, oh <Project Information> Okay. It's like okay we did what they asked for Or what the middle hand asked for and and we send a lot of data and then the one that should do the analyze. So now our our tool can only take one average point for every four square meters And So there's a problem with the data transmission or how to you say.	DQ DC DSU DS	ED PD
[01:18:40] SPK_3		
Yeah that too like okay. We maybe had a very high quality or very high precision, But their model couldn't use it anymore.	DQ DS DSU	PD
[01:18:50] SPK_1		

so maybe they just want to high level aggregate level and to compensate accuracy level I would say yeah		
[01:18:57] SPK_3		
Which which means that if you have actually, there was a small river or like my favorite one is just like 25 centimeters wide, but it's like going down like this. Yeah, very important for the water analyzes you do, but if you have one average and four square meter. Yeah, it will look like but on our it's like So it's a Yeah. So we have the we were just producing the data to someone else and we I think okay. This is strange but yeah		
[01:19:41] SPK_1		
So in that case then if there's a negative outcome, is it the People that make the decision based on this or you that produced the the basis for decision making that is responsible for the negative outcome. Or whose responsible for the negative outcome of that question. If there's a negative outcome.		
[01:20:04] SPK_3		
Interesting question. Yeah. <mark>it's actually the political who trust on the things that they decide and it's on on paper itself.</mark>		
[01:20:22] SPK_1		
So the people that actually make the final decision.		
[01:20:27] SPK_2		
Yeah. But they should be aware. Yeah.	ED	
[01:20:28] SPK_1		
and you are trying to make them aware of them. Okay issues? Yeah. Okay. Yeah. Okay, but interesting. Yeah,		
[01:20:37] SPK_3		
it's I don't think there's so many. Yeah. Actually, I think most of the analysis that is sent to this level higher level. It's not always done so mathematically and so on it's more. It's more like text-based and so on. It's not like they're looking at yeah. Yeah, but they very seldom look at every pixel at something or	ED	
[01:21:06] SPK_2		
Every pixel or like areas and so on like elections..		
[01:21:12] SPK_3		
Election there's a yeah,		
[01:21:14] SPK_2		
that's one when you change the valområde, valdistrikt What you say		
[01:21:22] SPK_1		
election areas.		
[01:21:24] SPK_2		
Yeah. Yeah, that's one thing that we do the analysis and everything and then it's like give them a map and then we write something and then it's up to the politicians. And yeah, basically, oh, this is it. I'm the same. Yes. Okay. Yeah that you always like yeah.		
[01:21:43] SPK_3		
but you can see from the states like depending on the district and so on there may you can get a different outcome.		
[01:21:58] SPK_1		
So I got one more question left, but it will be quite broad. It's for your personal opinion because I feel like nowadays. We collect more and more data for our future like supporting the decision making So to what extent do you feel like those data can actually representing us because you mentioned something like they trying to like		

maybe collect data for well-being and something that I don't know how to quantify it. So do you think that actually data? Is simplifying our daily life if we are basing on data to make decision for our future urban planning on City Planning? Like data, we use it like for simulation and then have analysis and then have three choices and afterwards they based on those data and just choose between the three choices. So it seems like okay, I only have three choices but actually people are irrational we can have a lot of choices. But now if you based on data you only have those limited choices for the future. So what do you think about it? Like the data driven urban planning, I would say.		
[01:23:11] SPK_3		
I think Okay, understand what you mean in one way you can go that Way. But in the other way what you're doing now when you using AI machine learning and so on before there was someone that by hand was like doing and they can do two maybe three use cases. Now we can produce hundred thousands of these cases And then we can just take some parameters and say okay, we throw 95% of them And we still have 10000 times more use cases that we fast can analyze and so it's not that we have less way to go but it's the problem is that which I have some mathematical as hard for that we go for things that is easy to measure and you mean that the things that is easy to measure and we tend to go in that direction. So Yeah. and the things that how to say money in one hand CO2 and the other hand. How should we..	DSU	PD
[01:24:30] SPK_1		
How should we strike the balance?		
[01:24:31] SPK_3		
Yeah, there was right there's quantified but daylight also hard to but other kind of pollution temperature. How do we move them or well-being? Okay. How do we so it's I think that. It's it's like always that now you can buy cherrypicking the right data, you can prove whatever you like. Which you could before too but you then it was easier to lie to or you can find them before you maybe can say my it is like this and then you just present it now. Now you have so much data and so instead of like saying just saying something it's like that you can share a pick data to prove your concept. does it answer your question in some way? Yeah, but	DSU	TCA
[01:25:28] SPK_1		
it's interesting to understand different people's perspective because we are this question to everyone.		
[01:25:35] SPK_1		
I just wanted. To make sure I don't know. I think we got answers too. But there's one that I just wanted to		
[01:25: 45] SPK_2		
We need to we need to find other room.		
[01:25:50] SPK_3		
we can sit at the table outside...		
a short break to change location		
[01:25:59] SPK_3		
I don't think we should talk about the system with you showed us. We maybe we should talk about our future goal project in. it's more Interesting to see that goal or		
[01:26:16] SPK_1		
So what are the goals that you are aiming for		
[01:26:19] SPK_3		

our goals is that the decision makers and we that works here <u>should have a common picture</u>. <u>Yeah, so or have the possibility to get the same picture because now all the people has different pictures about the future</u>. They think okay and		TCA MD
[01:26:40] SPK_2		
the present.		
[01:26:41] SPK_3		
and the present too. Yeah. Yeah, that's true and the history too maybe but a little bit like say, ah, we should build this here and they don't know that 20 meters away. There's a decision and plans that there should be something else. So yeah, you really have to know that to take a good decision there and if just 50% of the people know that here will be something at one place the decision on the next lot, or the next place is based on different grounds or thoughts and so on. So then it's a it's a bit for getting a more one picture and also that we will try to make people make it easier for people to get the right the same information because some people has a lot of information. And it's like yeah, it's yeah, you have it here in your bag here. So it's but I can't read it because you have it on paper in your bags. If we have it at the same place. Everyone has a possibility to get the same data And so on so so have a more harmonized of common ground to stand on yeah, so that's I think is the main goal for all things to you	BD DSU	MD
[01:28:15] SPK_2		
To be aware of what decisions will end up If you have the same picture.	DSU	TCA
[01:28:22] SPK_3		
Yeah, and probably the the it should be easy to get all the data and maybe also by mistake find data. So to say so I do this now by with the what's that? Sewage? enskilda avlopp		
[01:28:42] SPK_1		
The sewage. Yeah,		
[01:28:43] SPK_3		
so private sewage things so I mean you maybe don't think that you should think about this in one question, but if you see that oh This is an area where? Almost all has but three people don't have this. Okay, so then just say if you have a lot of this data easy to give to Reach and find so a little bit. I showed you the <Project Information>, for example, so the more say tell me everything about this area. So and if you then just spit out and you say oh here was something it's like there we collect for example information about fornlämning. What's that in English?		
[01:29:32] SPK_1		
I have no idea. I don't even know It means in Swedish		
[01:29:35] SPK_3		
so okay four thousand years ago. Someone died here in this buried here or here.		
[01:29:42] SPK_2		
archaeological.		
[01:29:45] SPK_3		
Archaeological part. So if they're exists some archaeological things we have that information from <Project Information>. So we have their data in our maps and we have analysis that go and get data from there. So if you say tell me what about here? So if they have any data about this area, we'll get that information easy. It's not like okay every time I have to call them up and say could you get out every paper you have on these and send them to me by post and I will read them and see if it's apply to my. So we have easier to get all the data on maybe more data than needed. Or then you think you need easy to get an also that they have a future. So that's the	DC DSU	PD

part that we're working with. Yeah, and then there's like the IoT part which may interesting you more but we're not so deep into that. We're more like we can we can provide you with a geography and the base for putting out the instead of you just getting text file or text which says X Y Z coordinate we can put it on the map and visualize it there for you. But now they today we don't have so much or almost no IoT data. I know that. They're in full quote on like elderly people. They have some test things and so on and yeah, then there's of course. On the small IoT that you don't think about. solar panels the charger the EV calls all it's in one way a part of it but it's not so integrated. It's just small Islands now, so it's not like integrated to bigger picture. I don't think that they have like okay if the sun is shining charge your car. And this kind of things I don't know maybe but		
[01:32:00] SPK_1		
I know there's an application that shows when the electricity is cheaper. And that's I guess when it's windy or sunny. and then when you should do things that takes a lot of electricity, I guess that's something similar. But		
[01:32:18] SPK_3		
yeah. Yeah,		
[01:32:19] SPK_1		
that's a private. Yeah		
[01:32:20] SPK_3		
now but yeah, it's for everyone's use and		
[01:32:24] SPK_1		
but I would say IoT is really expensive you if you want to extend to a larger scale to the city or something. That's just my personal opinion.		
[01:32:35] SPK_3		
We're talking they're starting to talk about this kind of things now a day even here all this kind of how to move energy and so on. But yeah, it's a that there you can see IoT in a different way.		
[01:32:49] SPK_2		
But I thought like I spoke to <Name> this morning and they got <Project Information> . Yeah. Yeah icy if it's icy then should you use your..first you need to put salt on the road or is it is it good if we are? We need to go in a different part of the city with our work. Is it good to bike or since it's ice outside then you can make the decisions. So that's that's one thing there they are starting now. So that's and of course, we've got the data if like the priority if we could stand on the roads, we got like priority road one, two, three, and of course you can make decisions out of that as well. Maybe we should need to change the order because it's this part of the town is really icy today. So that's		
[01:33:59] SPK_3		
That's it's thoughts and sound that they looked at and they've been at workshops in I think <Location Information>. They were workshop on this but and today we don't have the anything to measure today, but this was also one thing about the light poles if they're they can be an infrastructure for putting up. Iot things to measure this kind of things so we can do this kind of things in the future. But Today, we don't have it. It's probably the most closest it's like is there a parking lot free in the garage and so on there's we have iot for a long time. Like there's 25 places free there. So but this was temperature in this way. It's it's thoughts about it and they starting to build the infrastructure. <Project Information>		
[01:35:57] SPK_1		
I think that's the all questions that I want to ask. Yeah, is there anything you want to ask us?		

[01:36:02] SPK_3		
I assume that you will send us the or invite us to the presentation. And yeah,		
[01:36:08] SPK_1		
we will send you guys our thesis.		
[01:36:13] SPK_3		
Yeah, I don't need the printed version.		
[01:36:17] SPK_2		
Who else have you been interview?		
[01:36:19] SPK_1		
We've interviewed. I don't know if can that be Identifiable is that sensitive information? You know, we've just interviewed people that I have projects within digital twins in city setting. Yeah, okay, and we've interviewed both We've tried to find both decision makers and Technical people. Yeah to get both perspective. So yeah, but I'm actually not sure if that we are.		
[01:36:48] SPK_2		
It's fine.		
[01:36:49] SPK_3		
Yeah, it will be in your metadata today. Otherwise, I will not trust in your reference someone. Yeah, because we're otherwise.. I will sit in the audience. Okay. This is not a doktors-arbete (phd thesis). Yeah.		
[01:37:09] SPK_1		
It's good for us. Thank you so much for today. And for your interesting perspective. Yeah, thank you so much for using English with me. It will be easier to use the Swedish to talk about this topic I guess.		

Appendix 9: Interview 7, P8

Theme	Colour Codes	Factor	Text Codes
Context	Yellow	/	/
Data Management	Blue	Data Collection	DC
		Data Storage & Use	DSU
		Data Sharing	DS
		Data Quality	DQ
Algorithmic Bias	Red	Data	BD
		People	BP
Principle of Discourse - Democracy Principle	/	Pragmatic Domain	PD
		Ethical Domain	ED
		Moral Domain	MD
		The Theory of Communicative Action	TCA

Participants	
Authors	SPK_1
P8	SPK_2

Transcript	2nd layer	3rd layer
00:00:10 SPK_1		
Okay, so would you want to start with telling us a bit about the digital Twin project you're working with?		
00:00:18 SPK_2		
Okay. So we are building a product. We have built the product which is based on computer vision and the Principle is that we put cameras or cameras are embedded in different vehicles, taxis, buses, garbage trucks, maybe even your own car. And the the camera is embedded and looks forward to the the same perspective as the driver. And then we can run AI and computer vision to extract information from those video streams as the car drives through the city. And so the principle is actually we can extract useful information that is valuable to help those responsible for managing the city to manage it in a better way. So we focus initially on features like road damage and traffic signs and parking spots and things that the camera can easily see, and then extract that information from the video streams using computer vision and AI we send that to the city manager and then the city manager then has	DC DSU DS	PD MD

greater insights into the assets and the environment that they're responsible for managing. So in that we're kind of digitizing the physical environment of the city so that then it can be presented as data to those responsible. They can then have that greater insight, better updates and then more effectively manage their their resources to respond to issues and problems.		
00:01:48 SPK_1		
Okay, So is there. It's mostly the environment and not. People, movements and traffic flows so.		
00:01:57 SPK_2		
We could do that as well. We do have the possibility to detect people and vehicles, but I guess where we focused initially is on the the, the infrastructure, if I could put it that way. I think that's where we've seen the biggest pain, if you like the managing the city, that's where they see the biggest problems, the biggest opportunity for improvement. If they have greater insights and better data about what's happening, they can do things more cost effective effectively. They can save time, they can save money, they can make the place safer, they can make nicer to live. And I think that's where the real focus has been. Over time, I think we can start the focus on those more dynamic elements, but I it requires the more dynamic elements, the more vehicles you need out there collecting it, any kind of meaningful representation. So with the static things, you can start with only a handful of vehicles and already get. Very far, very quickly.	DC DSU	PD MD
00:02:55 SPK_1		
So for the data collection part, because you mentioned something like, you will collect data and then put it in like extract useful information. So how do you decide which area to run those cameras to collect data? What's the criteria or something like that?		
00:03:14 SPK_2		
Yeah. So actually you talk to the guy who's responsible at the city for that area, So for example in <Location Information> I think, The roads are managed by a team of <Number> and each person within team has is they divided <Location Information> up into <Number> different zones and then it's their responsibility to go and be responsible for one of those different zones. So if we were, when we work with <Location Information>, one of those guys puts a camera on his vehicle and then drives around his area. So all the roads have been driven and then because the camera is collecting data all the time, he then has a survey of his roads. That he can then that's now digitized so he has a digital twin of the road damage, like this area that's then presented on the computer in a map and can carry out analysis and then make a decision about which roads he wants to repair First (unhearable mumbling).	DC DSU	PD
00:04:11 SPK_1		
So you use the digital twin to decide where it's needed, Kind of?		

00:04:17 SPK_2		
Yeah, exactly. So, so if you think about the the city managers problem. Team manager has a pot of money from the taxpayer and he has roads he has to make safe and nice for cars to drive on and he wants to know what is the best way for me to spend that money? Where can I put that money to work in the most effective way and that might be repairing a particular road. Today it might be saying actually another road I don't want to touch today, it's fine. That's not a good way to spend this money. I should wait for next year to do that. And yeah, basically iterating and carrying on from there.	DSU DC	PD MD
00:04:58 SPK_1		
Okay. Shall we go to some bias part Okay. So yeah, so basically. From which area to collect the data? It will be up to the city planner, if I understand that correctly.		
00:05:16 SPK_2		
Yeah, I mean I think then it also depends on who, we also work for example with a taxi fleet here in <Location Information> is collecting data, we put cameras on all the taxis and then they collect data for the for the city manager on behalf of the city, OK. In that case the taxi is just driving around where they wherever they go and this is just if they have driving in that zone or Taking you to <Location Information> or wherever they're going, they're also collecting data, might cover some of the area that the guy's responsible for. And so when he comes to do his area himself, maybe he only has to do half the roads because the taxi	DC	PD
00:05:54 SPK_1		
okay. So you have some way of making sure that all the roads are being covered in the end.		
00:05:59 SPK_2		
Yeah. And actually one of the sort of next steps for us is to. It's just somehow influence the taxi driver, which is is not too hard, you just give them some money but the ways to to do that too. But if you think about the the road engineer's perspective, if they pay the taxi driver just to drive all of the roads in their area, then that might be a better use of time, might be a more cost effective solution. Rather than the road engineer having to equip their car and spend their time driving around the area, they just pay a taxi driver to do it for them. And actually that taxi driver can do it many more times. They can instead of doing it once per year, you can do it once per month and that gives them a week in better insight.	DC BD	PD
00:06:43 SPK_1		
So seems like you guys were also looking at the data quality, how is well representing the like the entire scenario in the real world. So is there any industry standard that you guys are using to analyze data quality or something like that?		
00:07:02 SPK_2		

So, so we we live and die by data quality in the end. I mean we want these guys to trust what we're telling them. So if if a road engineer is going to make a decision on how to spend money, they want that data upon which they're basing that decision needs to be trustworthy. So from a quality perspective it's it's absolutely crucial. When you say standards, I I'm not sure if you mean do you mean standards associated with data quality or standards associated with what data we provide them I guess. The question is driven that we we have there are frameworks that already exist for assessing roads for example. So when A walks down the road and looks at them that then he he is classifying the road according to a particular framework and we do the same thing using the computer vision. But right. How well we do that in terms of let's see then is A is another question and there are no standard frameworks that we know of in that context. That's just us assessing ourselves. Basically, we capture untruth when we compare ourselves around.	DQ BP	PD ED MD
00:08:03 SPK_1		
Mm HM, I think I'm also looking at the data quality, like if there is any uneven distribution or how well it's representing the metadata and something like that.		
00:08:15 SPK_2		
Do you mean in the training data then?		
00:08:17 SPK_1		
Yeah, yeah, yeah.		
00:08:18 SPK_2		
Yeah, okay, I misunderstood then. Yeah, for sure so. So I mean if that as well is you live and die by because you need to have an evenly distributed data set to make training, that means you can detect all the things you want to detect. We have a lot of there is a lot of work behind the scenes. We curate our data training data sets to make sure that our depending on the requirements of the model that we're trying to develop that the data sets are. They have enough representative examples of each class, for example.	BD	MD
00:08:52 SPK_1		
Okay, because you mentioned something like you guys also using AI technology in the entire things. So what's I know is a bit broad question. So what's your current understanding of algorithmic bias?		
00:09:11 SPK_2		
Algorithmic bias. So, so I think, I mean I I think it's perhaps the the sort of corollary of your question of of not doing what we do which is that if in the algorithm you train it with an uneven data set and it will passed according to the training that given it. So for example if if you will have a network that wants to you know a deep learning network that wants to detect a pothole on the streets or a crack on the streets. And in your training set, you have 100 potholes and one crack. Carry out the training. It's going to be really good at doing potholes and hopeless at doing cracks.	BD	MD

And the bias therefore is going to be that showing. When it drives around, the guy drives around the city, we're going to see lots and lots of potholes and cracks and that is then going to cause him to invest heavily in pothole repairing solutions, which may or may not be the right thing to do. Probably wrong.		
00:10:14 SPK_1		
Yeah, yeah. We just want to know like your personal opinion about that because we are curious about each interview with opinions on this algorithmic bias.		
00:10:23 SPK_2		
I I guess you I gave you a very specific, yeah, context of what I'm doing about. Yeah, it's a more general question.		
00:10:30 SPK_1		
it was a good example. Is this something like this bias outcome? Is this something that you regularly work with to counteract or?		
00:10:40 SPK_2		
Yeah, absolutely. For sure it's it's, it's linked heavily to the data quality, right. I mean if if you if you have a an algorithm that is biased then the data quality is not.	BD DQ	MD
00:10:55 SPK_1		
Is.		
00:10:55 SPK_2		
This I said, if you have an algorithm that is biased, the data quality is not high.		
00:11:00 SPK_1		
OK, yeah, we think that the connection is lost.		
00:11:03 SPK_2		
Yeah.		
00:11:04 SPK_1		
OK, OK. OK. I actually got an interest. No, it's. I don't know whether it's interesting or not, but to what extent do you feel like it's actually OK to accept bias? Because I know that there is a lot of contradicting like conflict of interest or different people have different agenda like city planner or you as a developer side. So to what extent do you think the bias is actually acceptable in the system if it's built for a urban planning or for the city?		
00:11:37 SPK_2		

So I mean I it perhaps it's not the the way I guess I'm not sure I I would focus so heavily on bias I know it's for your research so I it's not but I mean what what we care about is if the data is trustworthy or not you know so so as I say to what extent does it matter. Well it it it matters if the data is not trustworthy they they can't make any meaningful decisions based on it because they. You're telling me there's 100 potholes and 100 cracks, but I because there's bias, it's not true. Sorry, 100 potholes 1 crack for the example because it's. We know it's biased. It's therefore not true. Therefore my decisions are are incorrect and I'm not achieving what I wanted to do which was cost savings and safety increases. So actually if bias is the cause of that and is it be crippling any kind of value proposition that we offer If yeah if. If bias is less predominant and it's.	BD	PD
00:12:37 SPK_1		
I'm just wondering will those city planner actually ask about how the algorithm works in the system or will they want to understand more the how the technology works to give out those analysis or?		
00:12:51 SPK_2		
I think they want to understand to a level of very skin deep. You know, they like the fact that it's an image, they understand it's an image and a software boxes. Mmhmm. Things from the image, but quite. And they understand as well that there's there's a training process where you're teaching software as opposed to hand coding it, but not much more than that.		
00:13:12 SPK_1		
OK, so for the data that collected, so can I ask who will be the owner of the data?		
00:13:23 SPK_2		
So the our customers we do like training data or the data that's collected. For.		
00:13:31 SPK_1		
OK, So there is different ownership for different kind of data?		
00:13:35 SPK_2		
Absolutely. So data that is then labeled and trained and used in our processes and that produces a model that can infer other data. So when we create the model we use these, it's quite expensive to curate and produce a training set which then makes the model do what we want it to. You then deploy the model in the customer use case and then the model is detecting potholes and cracks for the customer at the point the data that is produced so that the locations of the potholes and cracks that the customer then used to repair, that belongs to the customer. That pothole location and that crack location, that's the customer's data. <Company> retains a right of use to it actually, because it helps us in improving and refining our algorithm, but that part belongs to the customer. Any data that is we use to train belongs to the <Company>.	DSU	PD

00:14:33 SPK_1		
OK, understand. So the time, OK, I'm just checking the time because I know that we've only got rid of it. Do you have any? I was. I forgot what I was thinking. Did you have the. Yeah. Did you think of anything right now? OK, so you've measured something like using the cameras to collect data. So will there be any possibility that we're actually capturing people's data in it or people movement data in it?		
00:15:10 SPK_2		
Absolutely. So that's one of the big challenges we've had as an organization is to recognize that we are by putting cameras in the public environment, we are at least potentially. Yeah, we need to manage personal data. So, so one of the when the camera is pointing forwards in the taxi taxis driving around, each image that is collected from the camera is then processed by <Company> software on the smartphone, out out on the edge, so to speak. And if useful is found a traffic sign or a pothole. If that information is then sent from the smartphone along with a copy of that picture. If nothing useful is found, so no traffic signs or no potholes, we destroy the image. We don't keep it in some way, but the only way any imagery leaves, which is what might contain personal data, is if something useful is found in it. That picture is then immediately anonymized. So if there is a picture of you standing next to a traffic sign detected. We will run some software that detects as a human being and then we will blur out the face the car. If we detected the license plate, we will blur out the license plates and then we delete the original image. So what we present to our customers is a detection of saying here is a pothole and here's a picture of it and any people or license plates will be blurred out and there's no longer any person.	DSU DS DC	ED
00:16:34 SPK_1		
Well, this is very intelligent. I was saying this is very intelligent. So do you think that actually? The policy is a bit hindering or maybe there is a risk that hindering that you guys developing this kind of intelligent		
00:16:51 SPK_2		
do you mean Gdpr?		
00:16:53 SPK_1		
Yeah, that's the first.		
00:16:55 SPK_2		
It is hugely hindering. It is a massive hindrance for us. It is a hindrance not because it's not a good rule and not because the intentions behind developing it were not good. It is a good rule and the intentions behind it were very, very good. The challenge is that people don't understand it and people actually, because they either don't understand it or misinterpret it, they react in completely the wrong ways. And actually what they intend to do then is react either, yeah, as I say the wrong way or with caution and they sit on their hands and don't do anything not to act and not to move forward as a result. So it we we've we've come a long way now, but initially it	DSU DC DS	ED MD TCA

was extremely hard to get. Particularly our customers who work in the public sector and are very worried about personal data and the of it's to move forward because whilst they saw a lot of benefits and a lot of gain, they also saw a lot of risk and and actually the the frustrating thing is that the risk is very, very low when we manage it very, very carefully. I think the the the misunderstanding about GDPR is that there is a, it is not private good, public bad or whatever simplified like that it's a much more nuanced shades of Grey and and so I think no nobody wants their personal data harvested behind their back so that some large <Country> corporate can make advertising I think everyone agrees to that but I I think everybody would be pretty happy to have some form of personal data perhaps aggregated to help make the road safe and. And so I think there's a, there is a sliding scale as long as people are informed and aware of the risks and do things with consent and have some kind of control and people are much more interested to engage and be involved. But I think the biggest challenge has been education. People just don't understand you the results they decide to forward and.

00:19:00 SPK_1		

So if there is a negative outcome in the project. Who is responsible for it?		

00:19:07 SPK_2		

What project are we talking about? Is this cuz we've developed a product and you're talking about if we engage with the city, who buys the product and there's a negative outcome?		

00:19:19 SPK_1		

Yeah.		

00:19:20 SPK_2		

OK. And what do you mean by a negative outcome? Whats?		

00:19:26 SPK_1		

For example, if there is, unfortunately there is data leakage. Then maybe there is certain like sensitive data that is speak to the public, maybe it is generate certain things like people aware of it. And then we consider this as a negative outcome because it's privacy, yeah, issues.		

00:19:48 SPK_2		

OK, so so if one of the images is captured and it misses the anonymizer and a picture presented with someone's face.		

00:19:55 SPK_1		

Yeah.		

00:19:55 SPK_2		

I I think. The answer is it depends I think if, if depending on who experiences that and talks to us about it. I think there is if someone wants to sort of really cry foul and cause problems for us, that could be really bad. But actually, but I think people are pretty reasonably minded and recognized that our intent is to do what we can to manage personal data and if that happens it's by exception and extremely unlikely and and we can manage it accordingly. So I mean if you think about, I mean you know most cars, everything about <Company> for example, most cars now have cameras on them running all the time, capture personal data all the time And in in future more and more car, Most cars coming off the production line have cameras looking out, capturing personal data all the time. And the way automotive manufacturers manage this is that they publish privacy policies and they say. If you, I think a car, a <Company> has captured a picture of you, you can contact <Company> and say, you know what, I think a vehicle captured my picture at this location at this time and you can request them to search their databases for such a picture and try and find it and delete it if you want to. And of course nobody does that, because people don't know about it, but they do offer the possibility to do that and that is a legal way for them to carry out.	DSU DS DC	MD ED
00:21:29 SPK_1		
Understand. OK, so currently this product is mostly focused on as I if I read it correctly and understand correctly in your website some of that is using for waste management for the city and also for traffic use as well as I remember correctly that.		
00:21:51 SPK_2		
So we work with waste management vehicles. Actually the application, they just carry the cameras, we're not. I mean the application is still around road management.		
00:22:01 SPK_1		
Ah, OK OK, so for this product, the data collected is for this purpose. So in the future, will there be any possibility that we use this data on other purpose I would say.		
00:22:13 SPK_2		
Yeah, for sure. I mean, that's the really exciting thing is that a camera is a very rich sensing modality, that there's a lot of information in an image. Today we've written software that it can extract particular features like traffic signs and potholes and road works. But tomorrow we're going to do parking spots and the next day we can do where the trees need to be cut back. And the day after that we can detect graffiti. And actually all of these things can be seen within an image. We just need to write the software to extract it and we start with particular features, develop the business and they do it.	DC	PD
00:22:50 SPK_1		

Nice. You have any questions so far? This time is almost yeah, I I was. So it's mostly infrastructure. Is there any connection or like something that could affect people somehow in this? That's maybe a bad question. Do you understand what?		
00:23:12 SPK_2		
I think, I think, I mean, the idea is that it's a really good effect on people, right? I mean the idea is that actually through this. Capturing this data, the city is maintained and managed in a much better way actually the city it is spent in a way that makes the road safer, that makes all of the issues responded to in a better and more cost effective way. And so the the intent here is to build cities that are much better run because what we've noticed is that the urbanization if you like people moving to cities despite COVID, it's it's increasing. The cities are just getting bigger and it's it's because people, you know, they seek out entertainment, entertainment, education and employment. And this is where these urban centers can provide most of that. But means that it's putting huge amounts of pressure on services and the infrastructure, particularly for older cities that were, you know, from medieval times and quickly in the <Country>, more modern cities are coping better, but they also were designed for particular population size and as they grow. The pressure increases and there's a need to do more with less, not least because climate changes. So our premise, our hypothesis, is it. The best way to do that is to get a better understanding of what's happening in the city, granular in in in your knowledge about what how the city is evolving. And based on that greater insight, you can make better decisions to save the time, the money, and make the place safe.	DSU DC	PD MD
00:24:45 SPK_1		
And do you have any insight on how this decision maker, decision makers decide where it's most important to start or is that solely their?		
00:24:57 SPK_2		
Well, I think their, their aspiration is the same is to say we want to build the city around people. I think previously they built cities around cars because transport in such a key thing and now there's a big push to say. Actually we want to remove cars from the equation as much as or at least diminish their effect on the city. In doing so we need to provide other transport options, so better buses, trains, bicycles, scooters, whatever it might be. But in in doing so we can also then having fewer cars and different transport options and we can then develop city around people as opposed to around the vehicles providing spaces for people to exist and. You know, better air and safer places and less crime and these kinds of things. So I think that that's the kind of backdrop of what they're all trying to achieve. And then using technologies like <Company>, they can get a better insight what's happening now, say if I now, if I understand much more accurately what's happening now, I understand what my goal is, what I'm trying to achieve. And then I can make plans for how best to get from where I am now to where I want to be in the most cost effective way.		PD MD
00:26:11 SPK_1		

So I know the time is almost there, but I want to ask a broad question. So now more maybe in the future that will be more and more data collect from the city like maybe some people data as well. So to what extent do you think that the the data can actually representing people like their human activities? Or maybe norms. I don't know cultures or things, because I know that now is kind of like using data to drive a lot of decision making for the city or for urban planning. So to what extent do you think data is actually representing people? Or will they actually have a risk of simplifying people or their human activities?		
00:26:57 SPK_2		
So so so can. So the question is is, is is. They're trying to capture data about infrastructure today, but they actually want to calculate capture data about human activity in future is that.		
00:27:10 SPK_1		
Yeah, yeah, yeah.		
00:27:11 SPK_2		
Yeah. So I think, I mean, I think so you mentioned before by digital twin, right. So there's constant producing some kind of digital representation of the city. You think about that as a concept you have. You could think of different layers in that digital twin and at the bottom you have the kind of built infrastructure, the kind of the buildings and the roads, and that changes at a certain rate. The next layer up, you have traffic signs and park benches and trees that is still almost static but changes much faster. Because trees grow and traffic signs fall over and then the next layer up you have much more dynamic data like vehicle flows and pedestrian flows and activity and things like that. And I think today we're we're at the base layer and we're moving up to the to the middle. And over time as as the the sensing and the technology and the number of vehicles and the number of IoT devices increases in the city, we will move higher up those those layers. And I would guess so I can imagine the ultimate objective is to have this digital representation. Have a living, breathing manifestation of what's happening out in the city almost on a second by second basis. And with that representation you then have possibilities not only to look back in time and say how have we, how have we, how has the city involved evolved and what state is it in and what do I need to do to repair it. We're also looking forward to carrying out sort of scenario analyses to say what if I changed the road in this way or change the bus route in that way. What might happen and carrying out sensitivity analysis of what can happen next and and improve decision making so that actually when you try to do something and invest in and change it in the city, you have that effect.		
00:28:55 SPK_1		
Okay, Thanks. So I think the time is almost, yeah, I think we got most of our questions answered as well. It was good. I'm.		
00:29:10 SPK_2		

So sorry again for being late. I really apologize for that and it was fun talking to you and if you need anything more give me a shout. What's the next steps is so you're you're gonna write up your thesis of yours.		
00:29:22 SPK_1		
Yeah, yeah. So we're gonna try to get everything down in the results and then do some analysis on the empirical data that we got. And that's that's pretty much the next step. Yeah. And. If you're interested, we can send you our final pieces as well.		
00:29:42 SPK_2		
Please do. I'd love to see it. Yeah. Thank you very much. It'll be really interesting to see.		
00:29:45 SPK_1		
Yeah. So before we close the call, do you have any question for us or do you think there is still something that we are not discussing that you want to talk about?		
00:29:55 SPK_2		
Nope, I think, I think you covered it. I just said I'm super keen to see the report when you're done. And yeah, but good luck with it.		
00:30:01 SPK_1		
Thank you so much. And also if if we remember something that we missed, is it OK to send emails and?		
00:30:09 SPK_2		
Yeah, of course. No, no worries still.		
00:30:12 SPK_1		
Yeah. Thank you. Thank you. So, OK.		
00:30:14 SPK_2		
Thanks for your time. See you later. Bye. Bye. Yeah.		
00:30:16 SPK_1		
Have a nice day. Bye, bye.		

Appendix 10: Interview 8, P9

Theme	Colour Codes	Factor	Text Codes
Context	<mark>Yellow</mark>	/	/
Data Management	Blue	Data Collection	DC
		Data Storage & Use	DSU
		Data Sharing	DS
		Data Quality	DQ
Algorithmic Bias	Red	Data	BD
		People	BP
Principle of Discourse - Democracy Principle	/	Pragmatic Domain	PD
		Ethical Domain	ED
		Moral Domain	MD
		The Theory of Communicative Action	TCA

Participants	
Authors	SPK_1
P9	SPK_2

Transcript	2nd layer	3rd layer
00:00:02 SPK_1		
Yeah, OK. So do you want to start by telling us a bit about your role, what you work with?		
00:00:11 SPK_2		
Yes. So yeah, my name is <Name> and I work for the city of <Location Information>. And that would be <Department>. I work a lot with preparing strategies for the increasing digitalization and automation of the transportation system. <Role>. <Company> approached us at the city and I said we have this idea, we want to develop self driving vehicles and we want to start a learning experience together with you and some other authorities as well and the universities. So that's what I'm working with and been doing for quite a long time now.		
00:01:15 SPK_1		

And the digital twin projects or projects you're working with?		
00:01:20 SPK_2		
Yes. And then it depends on your definition of the digital twin. I think I know that you sort of realized what I was doing while you were Googling <Project Information> and that is one project that we were part of and that project in specific was very much about how can we coinvestigate the use of a digital twin together the city the industry the university and so we had this pilot that we called <Project Information> it was about how can we make use our city digital twin to to learn how automated vehicles can interact with vehicles and the environment in a in a digital twin in a digital world. Yeah.		
00:02:20 SPK_1		
So sorry.		
00:02:22 SPK_2		
No, no. Let's go on.		
00:02:24 SPK_1		
What would you say are the overarching goals of this project?		
00:02:29 SPK_2		
Yeah, that specific project. It's about learning how to open up a sandbox environment in order for different organizations to to to work together and learn together how to make use of this.		PD
00:02:49 SPK_1		
Yeah, actually I got a question about the autonomous vehicles. What's is the intention on implementing this at first? Like what's the goal for it, like having autonomous vehicles at first in the city?		
00:03:05 SPK_2		
Yeah, the the overarching goals are that the automation is expected to bring safer streets, more sustainable environment, better services for more groups of people and then also that groups that today are perhaps underserved. They have a limited options of transportation. They will be able to get better options of transportations.		MD ED
00:03:33 SPK_1		
OK. Yeah. And who is participating in this project? Who's the stakeholders?		
00:03:42 SPK_2		

Oh, there are many in the digital twin <Company>. Yeah, there are many. University, Research Institute, city, maybe some governmental organization, industry, of course.		
00:04:03 SPK_1		
Okay and. Yeah. So when in this project, when different stakeholders have different goals, how do you reach a consensus like regarding what goals are more important or?		
00:04:20 SPK_2		
Yeah.		
00:04:21 SPK_1		
Yeah.		
00:04:23 SPK_2		
It it hasn't really been that much of a problem in that specific project because it's kind of an open innovation project. So, so there are the goals are not that specific, very detailed. It's more about learning, so I don't think that has been a problem.		
00:04:48 SPK_1		
So as you mentioned that there is a lot of partners participating in for example the <Project Information> projects. So how do you guys choose the participant? Are you like trying to choose a lot of people that will be affected or mostly the citizen or mostly from the business side or technical side?		
00:05:06 SPK_2		
In this project, I don't think that there were many citizens involved. It was more a business to government, to city cooperation.		TCA
00:05:18 SPK_1		
OK, so OK, so for the technology part of the Digital Twins, so it's more on the proof of concept instead of actually implementing in the city already.		
00:05:38 SPK_2		
You can say that we have and I always say that we have not one digital twin, we have many digital twins and it's about how to integrate them and make them even more provide add even more value to what they want to be used. <Project Information>. But then in in in the pilot I was involved then we investigated things and we understood new requirements that we would like to put on the digital twin and that is for instance. Ohh, how do you call it in?		PD
00:06:38 SPK_1		

English with she Is going to translate it		
00:06:39 SPK_2		
egenskaper like we wanted to know. Also we understood that it's not only that the façade of a building what it looks like with windows or colors and things and.		
00:06:54 SPK_1		
Texture development I don't.		
00:06:57 SPK_2		
Know.		
00:06:57 SPK_1		
Level details.		
00:06:59 SPK_2		
Yeah, yeah, level of details for sure. But more the properties of the materials are also of interest when you want to for instance investigate and simulate how how radio communication signals are transmitted in an urban environment. And and that has not yet been a prerequisite of the digital twin in <Location Information>. But in order to use it for for more those types of analysis then it's important to also know. More detailed information about the materials in the building, facades and so on. That was one finding for instance, and so so so in some extent for using for visualization or walking around in a three-dimensional 3 dimensional environment then. I would say the the city twin is fit for purpose but for other types of use then we found that we we need to develop the model in order to make it use for those types of studies.	DC	PD
00:08:01 SPK_1		
So, so far is the what I heard is about the infrastructure data like building. So is there other data that is being collected and put it in the digital twins, different kind of digital twins use cases.		
00:08:17 SPK_2		
We we have different other types of data that is digitized, but maybe then in in other twins so to say, and they are not always interconnected. And so when you want to combine data then you do the project by project and maybe data typically is not specified in order to be interchangeable between different models. That is typically not the case. So there are a lot of things to develop in order to make this twin really consists of many different types of data sources and then make them be able to interact a bit more smooth than is the case today.	DSU　　DS	
00:09:06 SPK_1		
Is there any people data or people movement data that is collected as well?		

00:09:13 SPK_2		
Not on a regular basis. We have done and people movement than I think of people walking or so yeah we have some data related to that that we collected in in one specific project. But we don't do that on a regular basis because that has also a lot to do about the intentions of GDPR and and so on. That's really personal information.	DC	PD ED
00:09:43 SPK_1		
Yeah, understandable. Yeah. OK. So how do you decide where this digital twin should be applied? Is it the whole city or?		
00:10:01 SPK_2		
Yeah, yeah, yeah. The goal is for it to to to that, to symbolize all of the city. Yeah. But of course, it started in in, in the very central parts of the city. So it's more detail and it's got more information in the central part, but the goal is for it to be able to, it will represent the whole of the city <Location Information>. Mmhmm.	BD BP	PD
00:10:28 SPK_1		
Could there be issues with? Like if the the center is better represented than the outskirts. How do you?		
00:10:37 SPK_2		
Yes, yes, I think so. And that is typically the case I would say and I think, but the I think people know this and people realize this and the goal is for the twin to to comprehend all of <Location Information>. To comprise all of <Location Information>	DC	PD
00:11:02 SPK_1		
This is something you are actively working.		
00:11:04 SPK_2		
Yeah, yes. But of course, typically there's always more interest in in, in, in, in the city parts, so to say, mmhmm, from industry, maybe from the research as well sometimes.	BP	PD
00:11:20 SPK_1		
Why is? Is there a specific reason for that or.		
00:11:23 SPK_2		
Probably because there are. More business opportunities, more people visiting and so on.	BP	PD
00:11:34 SPK_1		

So tourism as well?		
00:11:37 SPK_2		
Yeah, Tourism, probably, yeah.		
00:11:41 SPK_1		
OK. And bias. Yes. Yeah. Do you wanna? OK, Yeah. So to a broad question and. What's your current understanding of the algorithmic bias, uh?		
00:12:03 SPK_2		
AHuh [sounds a bit excited], I've never heard of that expression, but I could imagine and of course, yeah, yeah, yeah. If you have algorithms that usually have a output will not defined frame, so to say, then it will be biased the output of course.	BD	ED
00:12:22 SPK_1		
Yeah.		
00:12:22 SPK_2		
Yeah, yes, that. I would understand it would be the meaning of that, yeah.		
00:12:27 SPK_1		
Yeah. So even though you didn't come across the the term, is this something that you that's up for discussion in in the projects like this concept of bias?		
00:12:41 SPK_2		
In and not to my knowledge. Not to my knowledge.		
00:12:46 SPK_1		
And when collecting data in general is.		
00:12:51 SPK_2		
Yes, that I think people are a bit aware of. How, How you collect data and what type of data and what that might represent. And that that is important. I think that is important. And yeah, yeah. And I think many of my colleagues reflect on that.	DC DSU	ED
00:13:11 SPK_1		
Mmhmm, so is there any discussion or any company standard or industry standard towards the data quality, especially towards those data collected?		
00:13:23 SPK_2		
Oh, like.		

00:13:25 SPK_1		
For example, how to manage the metadata or how they yeah. Or maybe discuss how the how good is the data quality is.		
00:13:35 SPK_2		
Data quality, I think many people and in the industry are interested in and I think that is important and we have realized that, and we have. In some projects I work very much with how how should we put in place digital regulation, digital traffic regulations in order to be able to communicate with people and machines like self driving vehicles. And usually when you start developing and talk to other stakeholders, first focus is on APIs and then maybe data models. But then after a while, and we have come to this, but I would say not that many still yet on the European scene, in that you also need to work a lot to quality ashore your existing work process in order to feed this information system, in order to ensure high quality of the data, Because if you don't do that then you don't really know. What, what the data is representing and what you can make use of that. And I think that is still in the sort of discovery phase when it comes to because nowadays vehicles generate a lot of data. It's called in vehicle data and still some people think that yeah, we have this data and then we can make use of it in this, this or that way, but. Maybe that is not always the case, because maybe the data quality is not that good. Maybe the data doesn't really represent what you thought it would represent, it represent something else.	DQ DSU DC	ED MD
00:15:30 SPK_1		
So, so far the data are only collected for certain purpose and then it's not going to be reused in other area. So far or maybe in the future there is a plan for it.		
00:15:44 SPK_2		
And that is also a good point. We have come to that conclusion that we need to manage data in a much better way than we have done in history, and that is for economic purposes, but also for legal purposes and so on, because sometimes, and that is also the case in some European legislation, that one data set and, for instance, related to Traffic Safety. If if the use is to improve Traffic Safety, the data set could be retrieved and used for free. But if you have the exact same data set and you use it for another purpose that is not related to Traffic Safety for instance, and then you would have to pay a fee in order to use it. And so. So there there are so many new things that you need many new things that you need to think about when using data. And also some parts of your business, they might have procured or collect the data and they make use of data. And then another part of your organization wants to get hold of the exact same data sets for for their purpose, but they are not aware of this part of the organization. So then they maybe buy it or collect it or procure it. Once more. But we already had the data and that is not efficient. That's economically inefficient. Yeah. So, So this is very important and this is something that we are in the process of developing in our, in our city A.	DSU DS DC	PD ED
00:17:28 SPK_1		

Method or a business process too?		
00:17:32 SPK_2		
Yeah, yeah. So actually very much when it comes to connected and automated vehicles that I'm working with. Is that there are many new services that are needed to be put in place and those services are based on the exchange of data between different stakeholders and then you really. So sort of managing your data is becoming a new responsibility in your organization and that has not been the case previously. We are very much in the history of providing. Infrastructure for use in reality in real life to people. And if we and if we also have some data set available, then then it's just an extra deliverable, so to say. But in the future we must focus on delivering the infrastructure or the user available infrastructure both in real life and in the digital world at the same time. And then we need to change our mindset and we need to change our workflows. And that's what I'm working with trying to achieve this in our organization.	DSU	ED PD
00:18:42 SPK_1		
This I just wanted to you said that there's some new regulation, right? Do you know that where we can find information about that?		
00:18:52 SPK_2		
Oh, there are so many when it comes to data and the European Commission. There are so many legislations. One is the DATA Act for instance, and then in my working field it's the C or the ITS directive and inside the directive you have for instance RTI, delegated regulation, real time traffic information regulation that will impact us.	DQ	ED
00:19:27 SPK_1		
So this responsibility. It stems from regulations or yes, is there any other reason? Yeah.		
00:19:36 SPK_2		
Yes. And then of course you have GDPR, which is horizontal and the Data Act It's also sort of a horizontal and PSI directive and there are so many legal initiatives.	DQ	PD
00:19:54 SPK_1		
So maybe it's a kind of broad question as well. To what extent do you think that actually policy is hindering the growth of digital twins or using data, something like that?		
00:20:08 SPK_2		
Yeah, yeah. Policy or legal framework, I would say more because it's so complex and sometimes. People, people in general, they don't want to, they want to follow the rules. But if it's complex, then you might get scared and then you might ending up doing nothing instead of doing something.	DSU	PD ED MD

00:20:34 SPK_1		
Understand that we've heard that before.		
00:20:40 SPK_2		
I would assume so.		
00:20:42 SPK_1		
That's interesting, yeah. OK, so for the data collected, may I know that who will own the data? Like the ultimate ownership for the data ?		
00:20:55 SPK_2		
I think that's different in different cases and I would, yeah. And that you have to be very aware of. And that's also one reason why managing data is so important, you know to you have to really understand and know the terms and conditions for using the data who's owning it and in what ways, under what terms can you use it and so on.	DSU DS	
00:21:21 SPK_1		
Okay. So now that you're trying to re or you say that you are trying to form an organization that works with this new way of managing a city. Kind of.		
00:21:41 SPK_2		
And how I I try to influence my organization and take steps and try to to make our organization learn, parts of my organization to learn and I understand the future requirements that the people the the future expectations for the people that we are serving.		
00:22:01 SPK_1		
OK, so it's kind of research and development mindset.		
00:22:07 SPK_2		
Yeah, yeah, Development, yeah.		
00:22:11 SPK_1		
Yeah, yeah. Do you have, I feel like, yeah, it's nice that we're almost cover all our questions a, Yeah. Yeah. Practise makes Perfect I guess. Yeah. Maybe this one. Mmhmm. Mmhmm. Yeah, do you wanna? That's kind of your. OK, so maybe we need to take an example because everyone when I ask this question, it's really hard to understand because I think that there's a term that we use in the literature calling data-driven urbanism. It means that actually we're using a lot of data to make decision for our real life. So I remember there's one of the use case in the, I think it's <Project Information> like trying to use digital twins on a use case that when there is heavy rain, they are trying to simulate the different kind of situation how people will behave when there is heavy rain. Maybe they will go into some shelters or or going		

some different kind of direction. So when digital twins doing a lot of simulation, it seems like they are actually nudging people. To different kind of solution or different choice, do you think or to what extent do you think that actually data or digital twins is actually affecting people choices that in their real life or their freedom or liberty or something? Yeah, yeah.		
00:23:52 SPK_2		
I understand that.		
00:23:53 SPK_1		
Yeah.		
00:23:56 SPK_2		
Yeah, but that is already happening right now. So can.		
00:24:02 SPK_1		
Elaborate More like, what do you think that is happening right now?		
00:24:06 SPK_2		
Yeah, of course. If you run, if you're a sophisticated owner of an ice cream bar, then you look at the weather forecast and then you see, oh, it's going to rain though, then I understand I don't need to have that many people in my ice cream cafe shop because they won't come here and buy ice cream and so on. Predictive using data sources for predicting how to run your business. And then if you run <Restaurant> and you know it's going to be sunny weather than you know that people are not going to come to you, they're going to somewhere else. And yeah, with we don't do it that much in the city. But of course we know that if the weather is sunny, then maybe we need to have a more detailed look at the garbage bins in <The Park> or other recreational areas and so on. Of course we know that. And we know that if weather is going to be bad, we know that we're in the winter time. We know we have to prepare and do maybe. Do a better winter maintenance work we know if we know it's going below, for instance, and we have developed methods when we use data for doing that predictive maintenance and so to say. But we don't use in general data, so much for nudging people's behavior. In some cases we do because we have objective, we have goals that we maybe want people to move or or use their bicycle more for instance. And then of course if you go to our <Platform> and then you, you get suggestions and for some reasons walking and the biking are top choices and that's sort of nudging from and that of course can be debated whether and how much And a government or local government should nudge or influence their residents. But we do it in some cases, not that many cases, but of course businesses, I would assume, do that.	DSU	MD PD
00:26:34 SPK_1		
So you mentioned that there is a lot of partners that participate in the Digital Twins or different kind of Digital Twins projects. So should be there is some like. Maybe politician or actually people that making decision for the city planning. So when you		

guys having a discussion, do they have any priority on which kind of goals that is more important in planning those digital twins usage like maybe sustainability or? Economic growth or something like that. Is there any discussion about that? Like what kind of goals will be more important than the others? Or more on their consideration to implement digital twins?		
00:27:19 SPK_2		
No goals specifically related to the use of a digital twin? No, I'm not aware of that.		
00:27:27 SPK_1		
OK, so how people comes up those? Like because I think you guys want to implement digital twins for certain use case or actually put in use in certain yeah reality. So how do people's come up those requirement?		
00:27:44 SPK_2		
Yeah, that could be, that could be research driven, it could be industry driven, could be and I know that some usage of the digital twin is driven by the, That <u>we would like to have a more interactive discussion with citizens, for instance, so citizens can use the digital twin as a billboard and post suggestions and stuff like that. It's a new way of interacting with citizens.</u>		TCA
00:28:16 SPK_1		
Is that?		
00:28:16 SPK_2		
<u>With the method of communication.</u>		TCA
00:28:18 SPK_1		
OK, if they're in use already or it's just the idea, OK.		
00:28:22 SPK_2		
Yeah, <u>it's in use.</u> <Project Information>.		
00:28:29 SPK_1		
OK. Yeah, we'll Have a look at it,		
00:28:30 SPK_2		
It's been around for a while, yeah.		
00:28:33 SPK_1		
It's OK. Yeah. We'll be very interested on that part. Yeah. Thanks. There's something that you want to add on. I was thinking since there's so many stakeholders, if		

something like if some project or some part of the project would end up? With a bad outcome somehow. Who is responsible for that then?		
00:29:02 SPK_2		
Depends what is a bad outcome. It it it's not that easy to answer to that in in a general way. I I I think no, it's not that easy to answer. And what is a bad outcome? Of course, if you break any legislative rules, if you break GDPR, then that's not good and then someone has someone is responsible for that. But then again, there are other types of questions that are more of ethical questions, and I think that There will be more discussions related to ethical topics in the future, like how much data should we specifically as a public authority collect and where should we really stay out of this? In what way should we use it? How much should we nudge, for instance?	DC DSU	ED MD
00:30:11 SPK_1		
Where would you say that this new or? Where is this incentive coming from?		
00:30:19 SPK_2		
Ultimately it should be a political discussion, but I don't see that to a large extent at the moment. But I would expect that to become more important in the future. I hope so. And there has been some national initiatives I know the government installed. They call it <Project Information> a committee for ethical discussions about the use of information technology. <Project Information>. Maybe you're already aware of that initiative. I haven't talked so much about from them in a couple of years, but a couple of years ago they talked really much about. Their their work and and that should be important discussion even on a local level, but I don't think that has evolved much, yeah.		
00:31:16 SPK_1		
And where are you seeing this trend coming from? Have you discussed it with colleagues or where?		
00:31:22 SPK_2		
Yeah, I have discussed it with colleagues. Yes. And there has been some national initiatives and not only in Sweden, I know in <Countries> and in other places as well. And so I think those questions are important in order to to understand that, give a little bit more clear directions of how, how, how much, how far should you go. And you can do quite a lot with data, yeah. And why should it stay away? And and it's it's, it's more complicated than just just GDPR. It's about your intentions and so on.	DC DSU	TCA MD ED
00:32:12 SPK_1		
So May.		
00:32:14 SPK_2		
That, yeah, I think that is important.		

00:32:16 SPK_1		
Okay so where else that for the data collected? Is that solely collected by your institution or actually some data is collected by the private companies as well and then share?		
00:32:27 SPK_2		
Yes, sometimes we procure data and then it's collected by industry for instance. Yeah. So I think that that is that has been a trend for a couple of years that that government or local authorities, public authorities, they collect less and less data using their own methods. Instead procure data. And there's also this discussion that is quite ongoing that is about how detailed data should you procure. Should you procure so-called raw data or should you procure rather aggregated data or or even the analysis or yeah.	DC DS DQ	
00:33:12 SPK_1		
So there is no real data involved so far. Real data, real time data.		
00:33:18 SPK_2		
In?		
00:33:20 SPK_1		
In the Digital twins.		
00:33:24 SPK_2		
In the sense that we have <Project Information>, yes, in some cases we use real time data. But in this <Project Information> digital Twin then we showcased how to integrate in the pilot I was responsible for this <Project Information>. Then we showcased how to integrate live data. We showed for instance the use of <Project Information>.	DSU	
00:33:51 SPK_1		
Yeah.		
00:33:52 SPK_2		
That was integrated in between, but that was a pilot. It's nothing that is open to anyone or to the public or so.		
00:34:04 SPK_1		
OK. So for the data collected, have you guys think about like to make it like a open data and share to everyone or what's the consideration of not doing that? Yeah.		
00:34:18 SPK_2		

Of course GDPR is very important when it comes to open data or not. Yes, I have not been involved so much in those discussions. It's more under and I gave you the name of <Name> for instance. So I would think that and. And maybe some data we have procured or bought data under certain terms and conditions and then we cannot really give them away and so on. And that once again points out the need for efficient data management of your sources and so on. Yeah, I think that needs to be in place and sorted out. Maybe some parts of the digital twin can be open and they are open today and that's about walking around and and and this city model, that is open. And then there's also a question about. I think many people have started to realize more and more about security as well, because when you combine multiple data sources, if if we share something and you combine it with other data sources then some. Foreign country could understand where some installations are that that are important for national security that they shouldn't have knowledge about and stuff like that. And so that is also something that is very important to think about I think.	DS DSU	ED MD PD
00:36:01 SPK_1		
I was thinking about this procurement of data. Is there any Process or work process for ensuring the quality of data there when you're not collecting it yourself or.		
00:36:18 SPK_2		
Too little, I would say, and that is something I'm working on. <Project Information> for European cities and that is an ongoing discussion as well. Because there are there are very few quality standards. There are many standards that have been developed or harmonized when it comes to APIs, but very few standards when it comes to how how to assure certain quality levels of the data sets. And I think that is something that needs further work that needs to be established in order to have efficient procurements.	DQ DSU	ED PD
00:36:58 SPK_1		
And that's something you're working on internally then or?		
00:37:01 SPK_2		
Yes, and externally for instance if you use in vehicle data and we procure you know travel time data is one type of data set that would procure and of course this set there is. How to say the information in the in the set may be very different if you procure data from that has been generated from only sports car vehicles or moped vehicles. Then you have different, completely different speed, range and also the coverage of data. Depending on what typically this data is generated through vehicle fleets could be commercial vehicle fleets and so on. And maybe they only operate in certain geographical they are more likely to be in some geographical areas than others. And that will also bias data as we discussed previously.	DC	PD
00:37:58 SPK_1		
Yeah.		

00:37:59 SPK_2		
So there are a lot of things that needs to be developed in order to ensure a <u>more efficient procurement of certain types of data sets.</u>	DC	
00:38:07 SPK_1		
Yes. So is it up for discussion now even though there's no standards?		
00:38:14 SPK_2		
For it, yes. <u>It's up for discussion.</u>		TCA
00:38:16 SPK_1		
So internally in your projects and.		
00:38:21 SPK_2		
<u>Yes, and <Project Information></u>		TCA
00:38:26 SPK_1		
To develop standards.		
00:38:30 SPK_2		
Yeah, yeah. And. At first that <u>we have a common understanding of the problem and in what direction we do we want to go and how do you spur harmonization?</u> Maybe you need to put that in some long term strategic plans for financing certain types of development projects on a European basis. And then you have to have different project initiatives that work with these questions. So we, I expect it to be a long journey. Yeah.		TCA
00:38:58 SPK_1		
Yeah. For example, if as you mentioned like maybe there is some biased data and then it go into analysis or simulation and it turns out that there is some kind of result that maybe possibly you guys think that OK, maybe it's wrong or it's not accurate or to some extent it's not accurate but still usable. So when people making decisions towards this kind of analysis and results, will they want to understand more about how's the analysis come out or is it come out because of the biased data, Will they want to understand more or actually they just trust the analysis or they solely not trusting the analysis? Is there any situation or something experience you can share about that I?		
00:39:45 SPK_2		
Think that is very different depending on who is involved in this. But yes, I know that we, at least in my organization, <u>we're a bit interested in how this data has been generated and we want to understand how representable this data is for the phenomenon that we are trying to understand. So we'll sometimes do quality checks.</u> For instance, if we procure in vehicle data, then we would like to know for	DQ DSU	ED MD

instance how representative is the fleet and we've done some some research in some projects related to that in order to increase our understanding.	DC	
00:40:37 SPK_1		
So if the data that procure is not really reaching your standard or so, what will you guys do afterwards? Are you going to procure more data or yeah, I just want to understand.		
00:40:48 SPK_2		
Yeah, yeah, yeah. Yeah. Depends on the situation and the use as well. Sometimes it's just OK then this is the way it is and it's usually not when you talk about data-driven and so on and. It's how to say the causality is not always that exact, and it doesn't have to be. Maybe it's good enough that you know that the parameter is in the order in between this and this, this limits, and then it's okay for you to use. Okay. It doesn't have to be this or it can be this, so so, but sometimes it could be that the data is so to say, useless. We have some experience. We had one research project when we wanted to collect trajectories of people walking around in open areas and so on and. In that project the data that we collected with that method that it was useless for the purpose that we wanted to use the data and of course we understood that and then we couldn't do what we wanted to do. So so so if data is if quality is too low, then maybe you cannot do what you intended to do with the data. And in some cases, data is maybe good enough, because yeah, it doesn't have to be this precise.	DSU DQ	MD
00:42:31 SPK_1		
Yeah. So it's a discussion each time.		
00:42:34 SPK_2		
Yeah yeah, depending on what you want to use data for.		
00:42:38 SPK_1		
Mmhmm, but who will be the one that making the decision like I want to use this data or I'm not going to use it?		
00:42:46 SPK_2		
Ohh the the responsible project manager of course.		
00:42:49 SPK_1		
OK. OK. So it's also from your institution?		
00:42:54 SPK_2		
As well, yeah. Yeah. Yeah. From my department, yes.		
00:42:58 SPK_1		

OK, understand. I think we already cover most of the questions. Yeah, yeah, I'm looking at them. I think we got answers to them in other questions as well.		
00:43:10 SPK_2		
Good, Good, Good.		
00:43:11 SPK_1		
Yeah, maybe just this one. Yeah. So are there any further limitations or concerns that you see that we didn't discuss yet in this? At this topic, it's. A big question. Yeah.		
00:43:25 SPK_2		
Yeah, not that, not that I come to think of right now, but I think you have a important and you already understand this about data quality and so on and biased data and what is data really representing?		
00:43:42 SPK_1		
Yeah, OK. So are there any questions you would wanna ask us or?		
00:43:51 SPK_2		
Yeah, the the topic of your what was it?		
00:44:00 SPK_1		
Our thesis.		
00:44:01 SPK_2		
Yeah, thesis already already. Do your Was it for university or was it for Was it the master thesis?		
00:44:11 SPK_1		
It's a master thesis, yes.		
00:44:13 SPK_2		
Yeah.		
00:44:15 SPK_1		
Yeah. So we, yeah, we have been doing that since. Or finding the topic and trying to understand it since a few months.		
00:44:27 SPK_2		

Yeah, so we have a like 4 1/2 months allowed for for your thesis, but you can work on it longer.		
00:44:36 SPK_1		
Yeah, yeah. We actually, we have, I think it's two months where we are supposed to work on it full time. Yeah, but we start doing some things before that as well. It's just not official thesis. Time, but yeah, yeah, so, but we are going to finish it by end of this month and submit it or have the defense at the first week of June, yeah.		
00:45:01 SPK_2		
Yeah, yeah.		
00:45:02 SPK_1		
So it will be 1.		
00:45:04 SPK_2		
one Month to go one month.		
00:45:07 SPK_1		
Yeah, so this is actually our last interview as well. So yeah, we're just gonna try to synthesize everything and.		
00:45:15 SPK_2		
Have you talked to any of my colleagues in the city?		
00:45:20 SPK_1		
<Name>		
00:45:22 SPK_2		
Yes.		
00:45:23 SPK_1		
Yes, we tried to get a hold of <Name> as well, but <they> unfortunately didn't have time		
00:45:31 SPK_1		
Seemed we've got multiple people said that we should interview <them> so.		
00:45:39 SPK_2		
Yeah, <Project Information>		

00:45:42 SPK_1		
Yeah, yeah. But he gave us some other tips as well. Yeah, I think so. Mmhmm. So I think we can call it for today. And then thank you so much for your comments, opinions, perspective, everything. We think that would be interesting. Yeah. So if you're interested, then after we finish the thesis, we can send it to you to have a look as well if you want.		
00:46:08 SPK_2		
Yes, please. I would be happy to receive a copy of your thesis, Yeah.		
00:46:12 SPK_1		
So and if I was thinking if we. Come up with any other questions. Is it OK to e-mail you?		
00:46:19 SPK_2		
Yes, you can do that.		
00:46:21 SPK_1		
Yeah. And if you have any other questions, you can e-mail us as well.		
00:46:25 SPK_2		
Yeah.		
00:46:27 SPK_1		
So thank you so much and have a nice day ahead. I know that is the first day after the holiday. Yeah. Yeah.		
00:46:32 SPK_2		
It is.		
00:46:34 SPK_1		
Yeah. Thank you. So, thank you. Yeah.		
00:46:38 SPK_2		
Yeah. I wish you the best of luck and I hope to see your thesis done in mid June somewhere, yeah.		
00:46:45 SPK_1		
Good luck with the project.		
00:46:47 SPK_2		

Yeah. Thank you. Bye, bye.		
00:46:50 SPK_1		
Bye.		

References

Akter, S., McCarthy, G., Sajib, S., Michael, K., Dwivedi, Y. K., D'Ambra, J. & Shen, K. N. (2021). Algorithmic Bias in Data-Driven Innovation in the Age of Ai, *International Journal of Information Management,* vol. 60, no. 102387**,** pp 1-13

Ananny, M. (2015). Toward an Ethics of Algorithms, *Science, Technology, & Human Values,* vol. 41, no. 1**,** pp 93-117

Ananny, M. & Crawford, K. (2018). Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability, *New Media & Society,* vol. 20, no. 3**,** pp 973-989

Azevedo Guedes, A., Carvalho Alvarenga, J., dos Santos Sgarbi Goulart, M., Rodriguez y Rodriguez, M. & Pereira Soares, C. (2018). Smart Cities: The Main Drivers for Increasing the Intelligence of Cities, *Sustainability,* vol. 10, no. 9**,** pp 1-19

Batty, M. (2018). Digital Twins, *Environment and Planning B: Urban Analytics and City Science,* vol. 45, no. 5**,** pp 817-820

Bohman, James & Rehg, W. (2017). *Jürgen Habermas* [Online]. Available online: https://plato.stanford.edu/archives/fall2017/entries/habermas/.

Bovens, M. (2007). Analysing and Assessing Accountability: A Conceptual Framework, *European Law Journal,* vol. 13, no. 4**,** pp 447-468

Brauneis, R. & Goodman, E. P. (2018). Algorithmic Transparency for the Smart City, *The Yale Journal of Law & Technology,* vol. 20, pp 103-176

Buhmann, A., Paßmann, J. & Fieseler, C. (2019). Managing Algorithmic Accountability: Balancing Reputational Concerns, Engagement Strategies, and the Potential of Rational Discourse, *Journal of Business Ethics,* vol. 163, no. 2**,** pp 265-280

Burrell, J. (2016). How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms, *Big Data & Society,* vol. 3, no. 1**,** pp 1-12

Carlsson, S. (2003). Critical Realism: A Way Forward in Is Research. *Proceedings of the Eleventh European Conference on Information Systems.*

Charitonidou, M. (2022). Urban Scale Digital Twins in Data-Driven Society: Challenging Digital Universalism in Urban Planning Decision-Making, *International Journal of Architectural Computing,* vol. 20, no. 2**,** pp 238-253

Copenhagen Solutions Lab. (2018). *The Road to Smart Parking* [Online]. Available online: https://cphsolutionslab.dk/news/the-road-to-smart-parking [Accessed 29 December 2022].

DeCew, J. (2018). *Privacy* [Online]. Available online:
https://plato.stanford.edu/entries/privacy/.

Dembski, F., Wössner, U., Letzgus, M., Ruddat, M. & Yamu, C. (2020). Urban Digital Twins
for Smart Cities and Citizens: The Case Study of Herrenberg, Germany,
Sustainability, vol. 12, no. 6**,** pp 1-17

Deren, L., Wenbo, Y. & Zhenfeng, S. (2021). Smart City Based on Digital Twins,
Computational Urban Science, vol. 1, no. 4**,** pp 1-11

DeVos, A., Dhabalia, A., Shen, H., Holstein, K. & Eslami, M. (2022). Toward User-Driven
Algorithm Auditing: Investigating Users' Strategies for Uncovering Harmful
Algorithmic Behavior. *CHI Conference on Human Factors in Computing Systems.*

Dubber, M. D., Pasquale, F., Das, S. & Goodman, E. P. (2020). Smart City Ethics. *The
Oxford Handbook of Ethics of Ai.* pp 822-839.

Dulaimi, A., Hamida, R., Naser, M. & Mawed, M. (2022). Digital Twin Solution
Implemented on Energy Hub to Foster Sustainable Smart Energy City, Case Study of
Sustainable Smart Energy Hub, *ISPRS Annals of the Photogrammetry, Remote
Sensing and Spatial Information Sciences,* vol. X-4/W3-2022, no. 41-48

European Commission. (2021). *Local Digital Twins: Forging the Cities of Tomorrow,
Shaping Europe's Digital Future* [Online]. Available online: https://digital-
strategy.ec.europa.eu/en/library/local-digital-twins-forging-cities-tomorrow [Accessed
28 December 2022].

Evangelou, T., Gkeli, M. & Potsiou, C. (2022). Building Digital Twins for Smart Cities: A
Case Study in Greece, *ISPRS Annals of the Photogrammetry, Remote Sensing and
Spatial Information Sciences,* vol. X-4/W2-2022, pp 61-68

Ferré-Bigorra, J., Casals, M. & Gangolells, M. (2022). The Adoption of Urban Digital Twins,
Cities, vol. 131, no. 103905**,** pp 1-12

Floridi, L. (1998). Information Ethics: On the Philosophical Foundation Ofcomputer Ethics,
Ethics and Inf. Technol., vol. 1, no. 1**,** pp 33–52

Gal, U., Hansen, S. & Lee, A. S. (2022). Research Perspectives: Toward Theoretical Rigor in
Ethical Analysis: The Case of Algorithmic Decision-Making Systems, *Journal of the
Association for Information Systems,* vol. 23, no. 6**,** pp 1634-1661

Hämäläinen, M. (2021). Urban Development with Dynamic Digital Twins in Helsinki City,
IET Smart Cities, vol. 3, no. 4**,** pp 201-210

Häußler, H. (2021). The Underlying Values of Data Ethics Frameworks: A Critical Analysis
of Discourses and Power Structures, *Libri,* vol. 71, no. 4**,** pp 307-319

Helbing, D., Fanitabasi, F., Giannotti, F., Hänggli, R., Hausladen, C. I., van den Hoven, J.,
Mahajan, S., Pedreschi, D. & Pournaras, E. (2021). Ethics of Smart Cities: Towards
Value-Sensitive Design and Co-Evolving City Life, *Sustainability,* vol. 13, no. 20**,** pp
1-15

Helbing, D. & S'anchez-Vaquerizo, J. A. (2022). Digital Twins: Potentials, Ethical Issues, and Limitations. *Ssrn Electronic Journal,* pp 1-22

Herath, H. M. K. K. M. B. & Mittal, M. (2022). Adoption of Artificial Intelligence in Smart Cities: A Comprehensive Review, *International Journal of Information Management Data Insights,* vol. 2, no. 1**,** pp 1-21

Huang, W., Zhang, Y. & Zeng, W. (2022). Development and Application of Digital Twin Technology for Integrated Regional Energy Systems in Smart Cities, *Sustainable Computing: Informatics and Systems,* vol. 36, no. 100781**,** pp 1-10

Intersoft Consulting Services AG. (2016). *General Data Protection Regulation Chapter 3 Rights of the Data Subject* [Online]. Available online: https://gdpr-info.eu/chapter-3/.

Johnson, D. G. (2022). Algorithmic Accountability in the Making, *Social Philosophy and Policy,* vol. 38, no. 2**,** pp 111-127

Juvenile Ehwi, R., Holmes, H., Maslova, S. & Burgess, G. (2022). The Ethical Underpinnings of Smart City Governance: Decision-Making in the Smart Cambridge Programme, Uk, *Urban Studies,* vol. 59, no. 14**,** pp 2968-2984

Kitchin, R. (2016). The Ethics of Smart Cities and Urban Science, *Philos Trans A Math Phys Eng Sci,* vol. 374, no. 2083**,** pp 1-15

Konaté, J., Zaraté, P., Gueye, A. & Camilleri, G. (2020). An Ontology for Collaborative Decision Making. *Group Decision and Negotiation: A Multidisciplinary Perspective.* pp 179-191.

Kordzadeh, N. & Ghasemaghaei, M. (2022). Algorithmic Bias: Review, Synthesis, and Future Research Directions, *European Journal of Information Systems,* vol. 31, no. 3**,** pp 388-409

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G. & Yu, H. (2017). Accountable Algorithms, *University of Pennsylvania Law Review,* vol. 165, no. 3**,** pp 633-705

Lee, L.-H., Braud, T., Hosio, S. & Hui, P. (2021). Towards Augmented Reality Driven Human-City Interaction: Current Research on Mobile Headsets and Future Challenges, *ACM Computing Surveys,* vol. 54, no. 8**,** pp 1-38

Leese, M. (2014). The New Profiling: Algorithms, Black Boxes, and the Failure of Anti-Discriminatory Safeguards in the European Union, *Security Dialogue,* vol. 45, no. 5**,** pp 494-511

Living in EU. (n.d.a). *Local Digital Twin* [Online]. Available online: https://living-in.eu/groups/solutions/local-digital-twin [Accessed 28 December 2022].

Madsen, A. K., Grundtvig, A. & Thorsen, S. (2022). Soft City Sensing: A Turn to Computational Humanities in Data-Driven Urbanism, *Cities,* vol. 126, no. 103671**,** pp 1-12

Mahmud, H., Islam, A. K. M. N., Ahmed, S. I. & Smolander, K. (2022). What Influences Algorithmic Decision-Making? A Systematic Literature Review on Algorithm Aversion, *Technological Forecasting and Social Change,* vol. 175, no. 121390**,** pp 1-26

Mingers, J. (2004). Real-Izing Information Systems: Critical Realism as an Underpinning Philosophy for Information Systems, *Information and Organization,* vol. 14, no. 2**,** pp 87-103

Mingers, J. (2011). Ethics and Or: Operationalising Discourse Ethics, *European Journal of Operational Research,* vol. 210, no. 1**,** pp 114-124

Mingers, J. (2015). Discourse Ethics and Critical Realist Ethics, *Journal of Critical Realism,* vol. 8, no. 2**,** pp 172-202

Mingers, J. & Walsham, G. (2010). Toward Ethical Information Systems: The Contribution of Discourse Ethics, *MIS Quarterly,* vol. 34, no. 4**,** pp 833-854

Moggi, S. & Dameri, R. P. (2021). Circular Business Model Evolution: Stakeholder Matters for a Self-Sufficient Ecosystem, *Business Strategy and the Environment,* vol. 30, no. 6**,** pp 2830-2842

Mohammadi, N. & Taylor, J., E. (2018). Smart City Digital Twins. IEEE Symposium Series on Computational Intelligence, SSCI 2017 - Proceedings, 2018.

Mohammed, I., Alshuwaikhat, H. & Adenle, Y. (2016). An Approach to Assess the Effectiveness of Smart Growth in Achieving Sustainable Development, *Sustainability,* vol. 8, no. 4**,** pp 1-22

Nochta, T., Wan, L., Schooling, J. M. & Parlikad, A. K. (2020). A Socio-Technical Perspective on Urban Analytics: The Case of City-Scale Digital Twins, *Journal of Urban Technology,* vol. 28, no. 1-2**,** pp 263-287

Oates, B. J. (2006). Researching Information Systems and Computing, Londoin: SAGE.

Orlikowski, W. J. (2016). Sociomaterial Practices: Exploring Technology at Work, *Organization Studies,* vol. 28, no. 9**,** pp 1435-1448

Oti-Sarpong, K., Bastidas, V., Nochta, T., Wan, L., Tang, J. & Schooling, J. (2022). A Social Construction of Technology View for Understanding the Delivery of City-Scale Digital Twins, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* vol. X-4/W3-2022, no. 2022**,** pp 205-212

Papyshev, G. & Yarime, M. (2021). Exploring City Digital Twins as Policy Tools: A Task-Based Approach to Generating Synthetic Data on Urban Mobility, *Data & Policy,* vol. 3, no. E16**,** pp 1-18

Patton, M. Q. (2014). Qualitative Research & Evaluation Methods: Integrating Theory and Practice, California: SAGE.

Petrova-Antonova, D. & Ilieva, S. (2021). Digital Twin Modeling of Smart Cities. *Human Interaction, Emerging Technologies and Future Applications Iii.* pp 384-390.

Pierce, P., Ricciardi, F. & Zardini, A. (2017). Smart Cities as Organizational Fields: A Framework for Mapping Sustainability-Enabling Configurations, *Sustainability,* vol. 9, no. 9, pp 1-21

Rai, A., Constantinides, P. & Sarker, S. (2019). Next Generation Digital Platforms: Toward Human-Ai Hybrids, *MIS Quarterly,* vol. 43, no. 1, pp iii-ix

Recker, J. (2013). Scientific Research in Information Systems: A Beginner's Guide, Heidelberg: Springer.

Richardson, S. (2021). The New Physicality of Data, *Business Information Review,* vol. 38, no. 2, pp 67-74

Saunders, M., Lewis, P. & Thornhill, A. (2016). Research Methods for Business Students, 7, Harlow: Pearson.

Schiavone, F., Appio, F. P., Mora, L. & Risitano, M. (2020). The Strategic, Organizational, and Entrepreneurial Evolution of Smart Cities, *International Entrepreneurship and Management Journal,* vol. 16, no. 4, pp 1155-1165

Shahat, E., Hyun, C. T. & Yeom, C. (2021). City Digital Twin Potentials: A Review and Research Agenda, *Sustainability,* vol. 13, no. 6, pp 1-20

Stahl, B. (2008a). Researching Ethics and Morality in Information Systems: Some Guiding Questions.

Stahl, B. C. (2008b). Discourses on Information Ethics: The Claim to Universality, *Ethics and Information Technology,* vol. 10, no. 2-3, pp 97-108

Stinson, C. (2022). Algorithms Are Not Neutral: Bias in Collaborative Filtering, *AI Ethics,* vol. 2, no. 4, pp 763-770

Sunstein, C. R. (2022). Governing by Algorithm? No Noise and (Potentially) Less Bias, *Duke Law Journal,* vol. 71, no. 6, pp 1175-1205

Sweden.se. (n.d.). *Equality in Sweden* [Online]. Available online: https://sweden.se/life/equality [Accessed 1 May 2023].

Tzachor, A., Sabri, S., Richards, C. E., Rajabifard, A. & Acuto, M. (2022). Potential and Limitations of Digital Twins to Achieve the Sustainable Development Goals, *Nature Sustainability,* vol. 5, no. 10, pp 822-829

Ugwudike, P. (2022). Predictive Algorithms in Justice Systems and the Limits of Tech-Reformism, *International Journal for Crime, Justice and Social Democracy,* vol. 11, no. 1, pp 85-99

Vallier, K. (2022). *Neoliberalism* [Online]. Available online: https://plato.stanford.edu/entries/neoliberalism/.

White, G., Zink, A., Codecá, L. & Clarke, S. (2021). A Digital Twin Smart City for Citizen Feedback, *Cities,* vol. 110, no. 103064

Wu, H.-N., Yin, L., Zhou, T. & Ye, S. (2017). City Smart-Growth Evaluation System. *IEEE International Conference on Smart Grid and Smart Cities.*

Yang, K. & Stoyanovich, J. (2017). Measuring Fairness in Ranked Outputs. *Proceedings of the 29th International Conference on Scientific and Statistical Database Management.*

Yigitcanlar, T., Kamruzzaman, M., Buys, L., Ioppolo, G., Sabatini-Marques, J., da Costa, E. M. & Yun, J. J. (2018). Understanding 'Smart Cities': Intertwining Development Drivers with Desired Outcomes in a Multidimensional Framework, *Cities,* vol. 81, no. 2018**,** pp 145-160

Zhang, J., Chen, C., Zhang, Y., Cui, Y., Han, P., Meng, N. & Xu, Y. (2022). The Framework and Practices of Digital Twin City. *2022 IEEE 12th International Conference on Electronics Information and Emergency Communication (ICEIEC).*

Zou, J. & Schiebinger, L. (2018). Design Ai So That Its Fair, *Nature,* vol. 559, no. 7714**,** pp 324-326