The Use of a Data-Driven Digital Twin of a Smart City: A case study of Ålesund, Norway

The objective of smart cities is to produce smart decisions that will address sustainability and climate change through big data and citizen engagement. However, to achieve this, governments must address two barriers to citizen engagement. First, concerns about data privacy, and second, the difficulty big data poses to nonexperts. Big data by its nature produces a wide range of varying insights; it is difficult for nonexperts to understand the situations these insights describe, much less determine, and communicate priorities and solutions.

We propose that high-quality, 3D graphical digital twins (GDTs) of cities can be used to create 4D visualisations of geolocalised time-series to address the barriers to citizen engagement. Such digital twins are increasingly available and affordable. The method proposed here was applied in a city council meeting in Ålesund, Norway. It uses off-the-shelf hardware and a game-engine and creates immersive environments to convey multivariate data in a data-agnostic manner. It combines geographical information system (GIS) data, building information modelling (BIM), descriptive and predictive demographics, and internet of things (IoT) data from traffic telemetry, meteorology, telecommunication, and power. We demonstrate the benefits of the approach in multiple case studies related to the impact of Covid-19 in cities. The main contribution of this paper is the presentation of a novel, smart city GDT framework. It is lightweight for the IT-infrastructure; scalable in geographical size; transferable to other locations; versatile about data sources; privacy compliant; presents high-quality and reliable data; and intuitive because it is interactive, leveraging current advances in metrology.

Motivation for Data-Driven Digital Twins of Smart Cities

Cities are currently facing major challenges: accelerating population concentration, increasing traffic congestion and air pollution, and climate change. Urbanization is occurring worldwide; for example, 70% of the European population was urban in 2018 and 80% will be by 2050. According to the World Health Organization, 4 to 8 million people die prematurely because of air pollution every year. Two thirds of the world's biggest cities are coastal and will be impacted by sea level rise and more frequent and devastating extreme weather. Cities are also complex systems of systems. The problems they face are systemic and cannot be solved solely by optimising or upgrading existing cyber-physical infrastructure.

Addressing the short-term and long-term issues cities face will require high-quality data. It also requires effective use of such data. While big data offers the kind of comprehensive picture effective decision-making requires, it poses considerable challenges, such as the following:

- Difficulty with accessing the data and identifying which data are pertinent to a particular issue.
- Difficulty presenting data such that nonexperts can understand it.

- Finding a balance between high spatial resolution data and data privacy to preserve citizens' trust in local authorities and conform to European law, which takes general data protection regulation (GDPR) very seriously.
- Finally, ensuring that data are reliable and accurate.

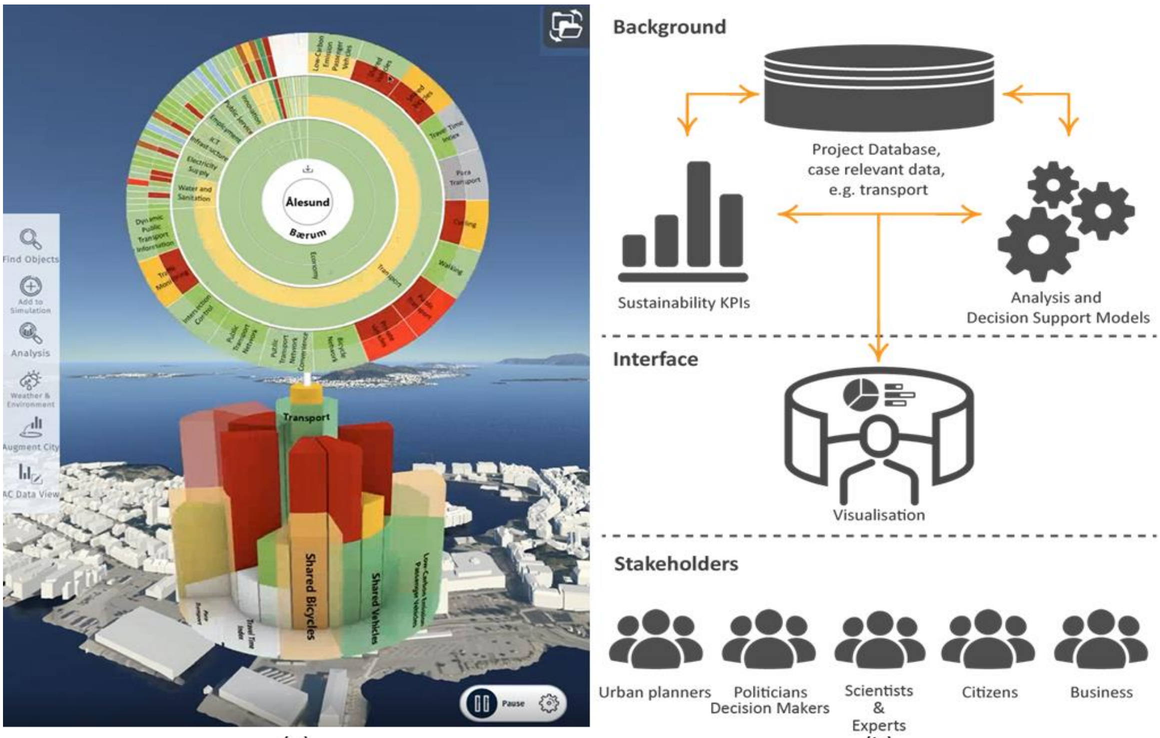
Researchers have drawn attention to the issues affecting the use of big data by local governments. For example, [1] calls for a smarter exploitation of useful data and critical interpretation of relevant measurements and assessment of their uncertainty.[2] lays out concerns related to privacy and accountability. Using a GDT, the authors in [3] resort to a semantic approach and data aggregation as a means of reducing data storage and transport, but do not address privacy. By contrast [4] proposes a geo-encryption approach to IoT sensor data stored on cloud servers as a means to address security, privacy, and cross-border legal issues. We argue in this paper that anonymising and aggregating data can create GDPR-compliant data to address these issues.
Many studies promote the use of IoT to foster understanding and improve various aspects of smart cities in real time or for long-term planning. For example,[5] demonstrates the utility of a hierarchical framework of microscopic simulation based on MATLAB and SUMO and real time events for optimizing traffic management. [6] shows that real-time data from IoT sensors and actuators, GIS, BIM, and meteorological information can be exploited to generate heating strategies that will find a balance between user comfort and power consumption from the building to the district levels.

Even though visualisations allow researchers to explore and set hypotheses, confirm and disseminate results, and interact with the public [7], most scientific literature fails to address visualisation. Smart city data often appear in the form of geo-referenced time series. Such data allow interpretation of priorities, decisions, and measurements in their spatiotemporal context. In [8], air quality data in a city is monitored without representation in a GDT of the city, thus failing to engage stakeholders. Moreover, [9] goes further by presenting a context-agnostic distributed air-monitoring system, thus decontextualizing the data and the insights context might provide. By contrast, the purposes, challenges, and advantages of representing big data in a GDT are elegantly summarized in [10], using data from three years of taxi rides in New York City. To create insight, the data must be contextualised with various data sources and spatio-temporal information. Furthermore, in an attempt of structure the domain data, GDTs of city frameworks and ecosystems have been developed by and for research projects, such as 3dcitydb or Fiware[1], but the semantic models they require prevent or hamper holistic digital twin approaches when adding data sources not fitting the relational data model. The approach presented in this paper is data-source agnostic: the system does not have a semantic model for presenting the data. Thus, the presenter must provide a narrative which makes sense of it.

The UN Sustainable Development Goals (SDGs) lay out issues city governments should be prioritizing as they commit resources to finding solutions to contemporary problems. They define sustainability as the balance between economic growth, social equality, and environment preservation and the key to the problems the cities face. The United Nations for Smart Sustainable Cities (U4SSC) initiative is one of a series of independent smart city initiatives that have mapped the SDGs to a set of 92 concrete key performance indicators (KPIs) from the three domains, summarized visually in a wheel chart in Figure 1 (a). Many

[1] http://www.fiware.com/

Norwegian cities have adopted them as a compass for smart and sustainable city-level decisions. Like the decisions they can trigger, the KPIs are related to geographical location and the underlying data, which often comes from sensors, and are thus the common denominator for the cases presented subsequently. Figure 1 (b) illustrates the interaction of the stakeholder with the visualisation of the KPIs and insights.



(a) (b)

Figure 1: Wheel of U4SSC KPIS in Ålesund, courtesy of AugmentCity (a) and Participative model in city council, courtesy of Ålesund Kommune (b)

The previous section emphases the need for using GDT to contextualize smart city big data. This section presents the architecture of such a digital twin. A data-driven digital twin framework was employed for this paper's examination of the impact of Covid-19 in Ålesund, Norway. Its layered platform architecture is presented in Figure 2 (a). The upper layer represents the graphical user interface. The simulation management allows the user to choose the data (from the catalogue) and the visualisation type (column, heatmap, etc.), to have time control (play, pause, slow-motion, fast forward, rewind, replay), and to navigate in the virtual 3D world with a joystick. The Unity-based visualisation module on the user interface layer can handle either a single display or a multi-projector dome for better immersion and concentration of users. The high frequency coordination between the visual clients and the core relies on a distributed gaming middleware. The core, implemented in Java, is the task scheduler responsible for controlling the scene and the time and for handling the requests from the simulation management and coordinating the data requests and reply payloads from third party data sources. An abstraction layer consisting of plugins, indicated in orange in Figure 2 (a), implements different application programming interfaces (APIs) and file formats. It also incorporates the gateways to data stored on third-party premises. This allows the platform to flexibly connect to RESTful APIs or comma separated value (CSV) files containing geolocalised time series.

The visualizations presented in this paper were performed on an affordable gaming gear, such as Dell Alienware Aurora R7 desktop gaming-PC, with an Intel i7 Processor and an NVIDIA RTX1080 card equipped with a 3D-Mouse for navigation. The distributed visualisation bus displayed in Figure 2 (a) was thus running on the local host with only one wide screen visual client.

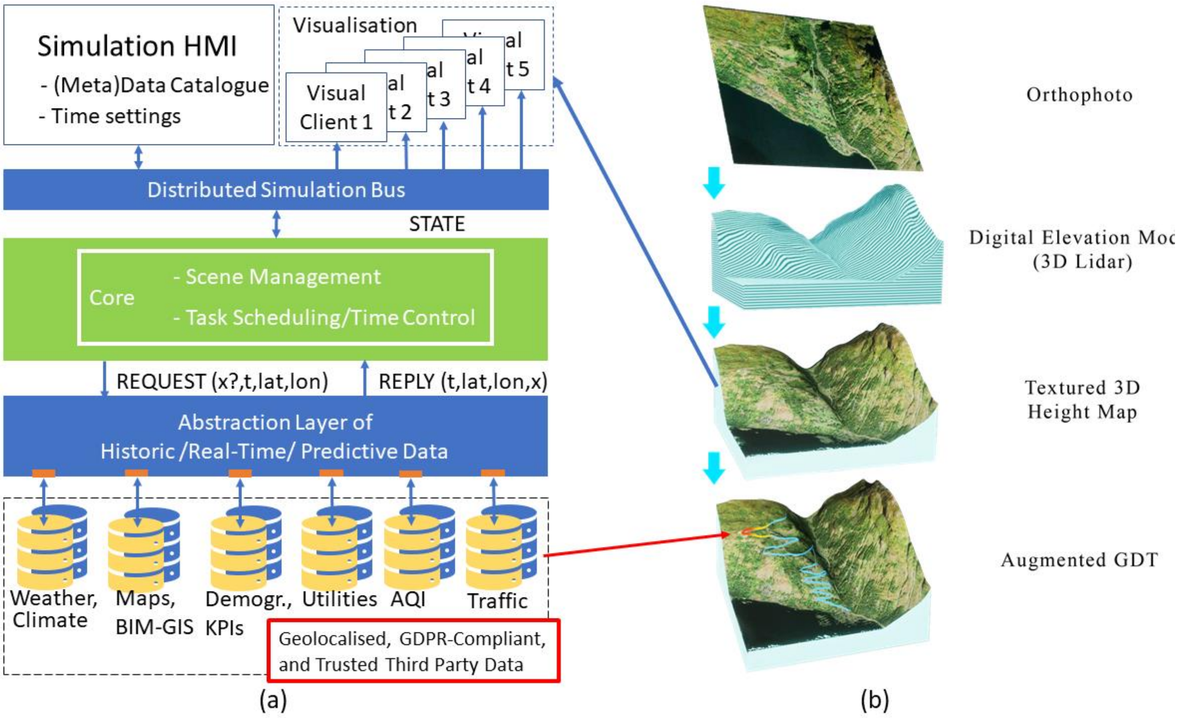


Figure 2: Architecture (a) and Graphical Digital Twin Method (b)

Graphical digital twin and visualisations

To place the data in spacio-temporal context, textured 3D digital elevation models (DEM) must be created. They are based on high-quality orthophotos are projected on high resolution public domain LIDAR measurements., as seen in Figure 2 (b) Once created, the 3D DEM is used to contextualize geolocalised time series, consisting of data points of which in Table 1 gives an example. The user can then navigate through space with a joystick and time by controlling the visualisation speed such as slow motion, hyperlapse, and stopping time on a snapshot. This allows stakeholders of different literacy to apprehend the visualisation at their respective speed. Inspecting data from different viewpoints in the 3D visualisation while navigating through the time dimension is called 4D Replay.

Table 1: Inductive Coil Data: Traffic Geolocalised Time Series

Time	Lat	Lon	Hourly Total
2020-03-31 14:00:00+01:00	62.471646	6.2149	**390**
2020-03-31 15:00:00+01:00	62.471646	6.2149	**454**
2020-03-31 16:00:00+01:00	62.471646	6.2149	**357**
2020-03-31 17:00:00+01:00	62.471646	6.2149	**297**
2020-03-31 18:00:00+01:00	62.471646	6.2149	**271**
2020-03-31 19:00:00+01:00	62.471646	6.2149	**251**
2020-03-31 20:00:00+01:00	62.471646	6.2149	**178**
2020-03-31 21:00:00+01:00	62.471646	6.2149	**113**
2020-03-31 22:00:00+01:00	62.471646	6.2149	**98**
2020-03-31 23:00:00+01:00	62.471646	6.2149	**38**

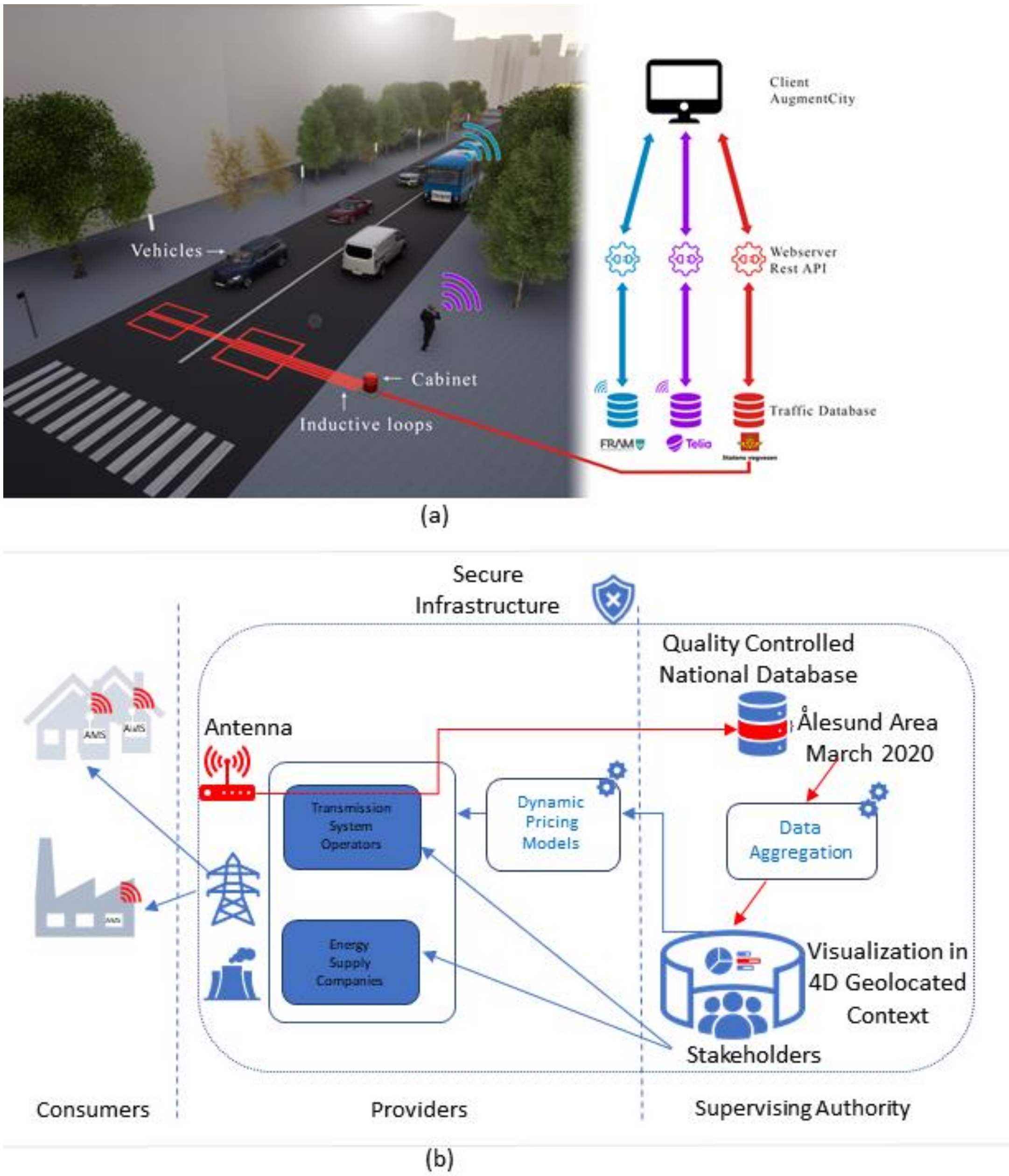


Figure 3 : Traffic Data Infrastructure, Courtesy of OSC (a), Electricity Data Infrastructure (b)

Table 2 : Overview of data

Type	Description	Relevance	Scalability	Privacy
BIM	Building and Road Information CAD, Timed BIM	Current and future city Contextualise energy consumption, infrastructure needs	Local and private data, international standard format	Occasionally critical
Energy & Water	Energy and Water Usage in Districts kWh/hour, m^3/day Geolocalised Time Series	Civil defence Current and future infrastructure needs	National database	Critical
Weather	Historic and Forecast Geolocalised Time Series	Contextualize outdoor activities Civil defence	Worldwide Service	Irrelevant
Air Quality	Historic and Forecast Geolocalised Time Series PM10, PM2.5, PM1, NOx, SO_4	Health and environmental consequences	National Service	Irrelevant
Traffic	Inductive Loop Data Geolocalised Time Series Vehicles/hour/day etc	Network utilisation of vehicle and bike infrastructure, emissions, congestion	National Service	Occasionally critical
Public Transport	Automatic Passenger Counting Per Bus, Ferry Line, and stop Passenger; Revenue	Monitoring and planning of infrastructure	Regional Service	Critical
Demographics	Historic and Forecast Geolocalised Time Series (age, sex, wealth, school pupils)	Contextualize and plan infrastructure needs: schools, roads, bus lines	National Service	Irrelevant
U4SSC KPIs	91 Sustainability KPIs	Identify priorities for sustainable planning	Worldwide Service	Irrelevant
Emergency Response	Fire and Ambulance Response Time Minutes to Destination Geolocalised Time Series	Identify areas with poor coverage Plan infrastructure	National Service	Critical
AIS	Automatic Identification System Geolocalised Time Series	Air quality correlations Traffic planning	Worldwide and National Service	Occasionally critical
Outdoor Activity	Geolocalised Time Series Outdoor Activities with Strava Outdoor Path	Contextualize outdoor activities Identify preferred routes	Worldwide Service	Occasionally critical

Case Studies

This section showcases the benefits of GDTs by illustrating data from the Covid-19 situation. Many data sources from local, national, and international providers were gathered and visualized. The study data consist of week 9 through 12, March 2020. Norway announced a soft lockdown at the end of week 10 and it took effect week 11. It is convenient to aggregate and contextualize data from heterogeneous public and private providers to a common basic geographical statistical unit, defined by the Norwegian central bureau of statistics. Table 2 summarizes the types of data accessible in the tool, and analyses their relevance for smart cities, their scalability to greater geographical units, and their respect of privacy.

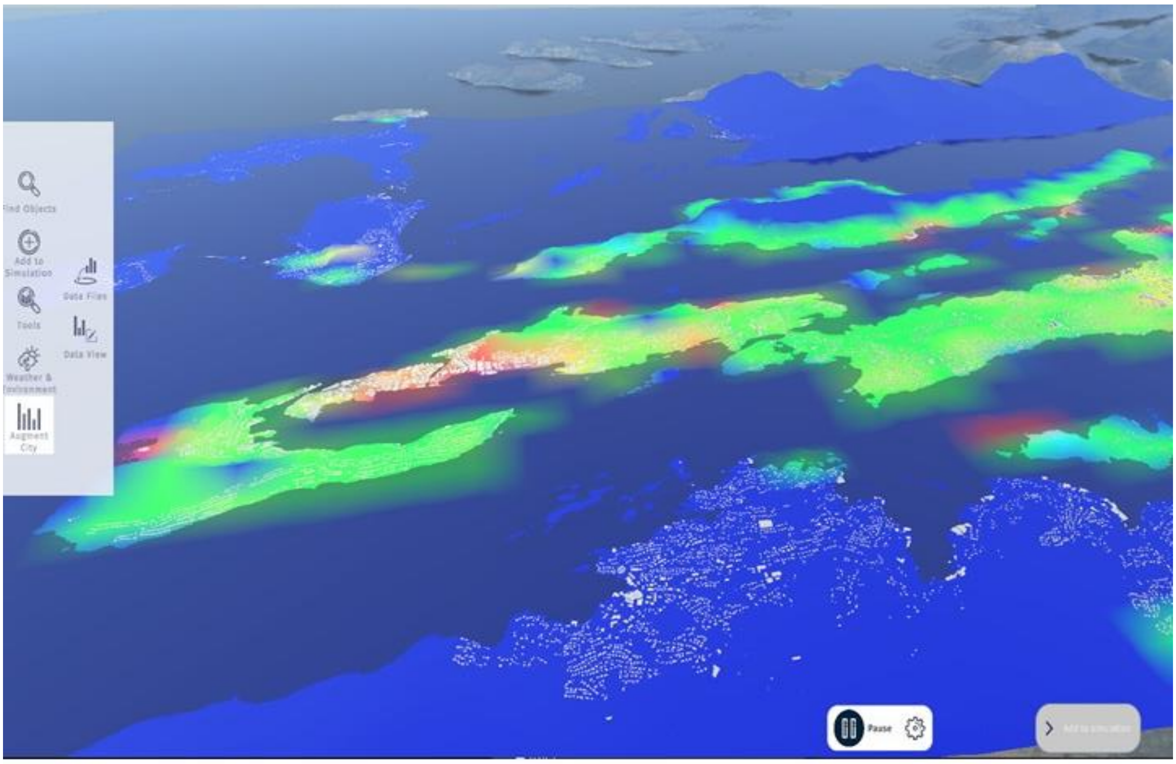
Firstly, Mobility is one of the major challenges affecting cities. Measuring and visualising the flow of persons and vehicles in a city makes it possible to understand mobility patterns in urban agglomerations. Mobility data can be gathered in many ways: IoT Data, cameras, GPS-data from mobile phones, manual counting and surveys, Bluetooth and Wi-Fi beacons, and cellular triangulation. Our case study utilises IoT data from the Norwegian Roads Authorities (SVV) and cellular data from the mobile telephony company Telia, and its Crowd Insights service.

Inductive coils permanently installed under the asphalt of the roads are connected to cabinets which detect changes in two coils' impedance due to the vicinity of metallic objects, as shown in

Figure 3 (a). These coils measure the velocity and length of vehicles in each road direction on the road above them. The cabinet forwards the traffic measurements to a central server, in open access via a Rest API. The measurements use a coarser aggregation: hourly average of both directions and by vehicle length or type (bike or vehicle).

Table 1 shows an excerpt of the processed data, the hourly bidirectional traffic at a measurement point, downloaded via the API. While Figure 5 (b) depicts a 2D representation of the time series at the measuring point, contextualising the time series provides more insights. Figure 4 (b) is a snapshot of the GDT displaying the hour-by-hour comparison of bidirectional traffic of the first 2 weeks to the last 2 weeks of March. The columns are placed at the position of the station in the GDT. The blue, red, and green columns represent, respectively, traffic before lockdown, during, and the difference between them. The interactive controls are visible in the picture. The window on the lower right-hand side is the time control and the window on the left controls the data and visualisation type selector. Aggregating the data on a weekly basis suggests a decrease of 30% of the traffic between the two first weeks and the two last weeks of March.

Secondly, Air pollution is a silent killer in cities around the world and is thus closely monitored. The historic air quality data curated by the Norwegian Air Research Institute (NILU), encompasses the hourly means for Particulate Matter smaller than 10µm, PM10, NOx, NO, and NO_2 in the city centre of Ålesund. Figure 5 (a) shows the timeseries for PM10. The Norwegian maximum regulated mean daily value is added to indicate that the threshold was reached only one day in March 2020. The value peaked in the first week; then level began to drop and the values for PM10 stabilized around 12 µg/m^3 before the lockdown was announced; see for comparison Figure 5 (b). We cannot infer causal relationships between air quality and the lockdown in this short period. This indicates that air pollution is a multivariable problem which is not easily and explained through a visualisation for such a small time-window.



(a)

(b)

Figure 4: Heatmap of energy usage one day March 2020 (a); Inductive Loop Traffic Measurements March 2020, Ålesund (b)

Thirdly, understanding the demographics of a city, its evolution over decades, and its circadian variations is critical for city planners. The demographic data presented in this case study was collected by the mobile telephony company Telia, using triangulated cell-phone position.

Figure 3 (a) sketches the architecture for gathering, storing, and presenting aggregated mobile user data. The data received are aggregated per hour and per basic unit. To ensure complete compliance with GDPR, a minimum group population of 5 persons was implemented. Figure 5 (d, e, and f) depict hourly scatterplots of the people in each basic geographical unit in the few days before (blue) and during (red) the weeks of lockdown in various districts. The data suggest a dampened circadian variation during the lockdown in the academic and residential districts, while the recreational area experienced more daily variation during the weekdays after the lockdown began.

Finally, the secure infrastructure supporting the gathering of private data and displaying them in a privacy-preserving manner is detailed in

Figure 3 (b). The transfer of the data from the consumers is handled by an advanced measuring and control system (AMS). It is installed at the premises and centralised in a nationwide datahub[2] delivered by the local energy provider Mørenett. The data can be visualized as a heatmap as in Figure 4 (a), or a time series as in Figure 5 (c). While the time series allows a better fragmentation of the data into aggregated residential or industrial commercial, thus identifying a reduction in consumption at industries in the two later weeks, the heatmap shows the data in a spatiotemporal context, with industrial district clearly identifiable with the red zones compared to the residential areas in green.

Discussion

The previous cases illustrate the usefulness of visualisation of measurements and draw our attention to different facets of smart city. One of the challenges of local government is the trust of their citizens, which is partly earned by guaranteeing their privacy and providing them with accurate information.

Privacy is achieved by means of GDPR compliance: aggregation, filtering, and fragmentation of the storage contribute to the anonymisation of data. No critical data leave the dedicated IT premises and only privacy-compliant data is exchanged with and stored on the 4D system. This introduces a necessary step, and while it might lead to latency, real-time information is not critical for the long-term planning decision process.

The information accuracy is addressed in Table 3, which lists the level of uncertainty for the measurements of the study. Firstly, the electricity company Mørenett reports that the accuracy of the AMS under scrutiny present an accuracy in the regulated range, with 0.2% for a 2% normative range. With a normative 2% restriction, the AMS accuracy seems very high, but when national annual electricity productions of European countries are in the 10 to 100 TWh range, this means a discrepancy of 200 to 2000 GWh, which converts to a yearly cost of millions of euros born either by producers or consumers. Secondly, inductive loops and SIM card data accuracy are not regulated by national directives in Norway, but their accuracies are necessary for daily traffic estimation and calibration of mobility simulation

[2] www.elhub.no

models as they often are deemed as “ground truth”. The accuracy estimations reported by the Swedish and Norwegian Road Authorities (Trafikverket and SVV) are appropriate. Finally, the *European Standard for gravimetric measurement method for the determination of the PM10 and PM2.5 mass concentration of suspended particulate matter* (EN 12341:2014) requires an uncertainty of maximum 25%. This margin seems very large, but the purpose of the directive is to compare measurements from a wide range of automatic and affordable apparatuses and to guarantee their widespread use throughout the European Union and associated countries. The uncertainty of 12.8% declared by the Norwegian Institute for Air Research (NILU) is within acceptable range of the standard. As a concluding remark, while understanding uncertainty is a sound scientific practice which ensures long term trust in national metrology services, conveying uncertainty to the public is challenging as it adds complexity to the visualisation and risks to water down the message to disseminate.

Table 3: Measurement Uncertainty

Measurement	**Uncertainty**	**National Regulation**	**Model**	**Frequency of Calibration**	**Source**
Energy Meter (kWh)	0.2%	Maximum 2% error	Kamstrup DK 8660		Mørenett
Inductive Loop (car/hour)	1 to 9%, 95% of the time	Not Applicable	Datarec7/410/500	Monthly Quality Assurance	SINTEF, SVV, NTNU
SIM-Card Localisation	5% 95% of the time	Not Applicable	Base Antenna and Machine Learning	At training time	Telia, Sensebit, Trafikverket report
Air Quality (PM10)	12.8%, 95% of the time	Maximum 25% uncertainty	TEOM 1400AB	Cleaned weekly, filter changed quarterly	NILU, Ålesund Municipality EN 12341:2014

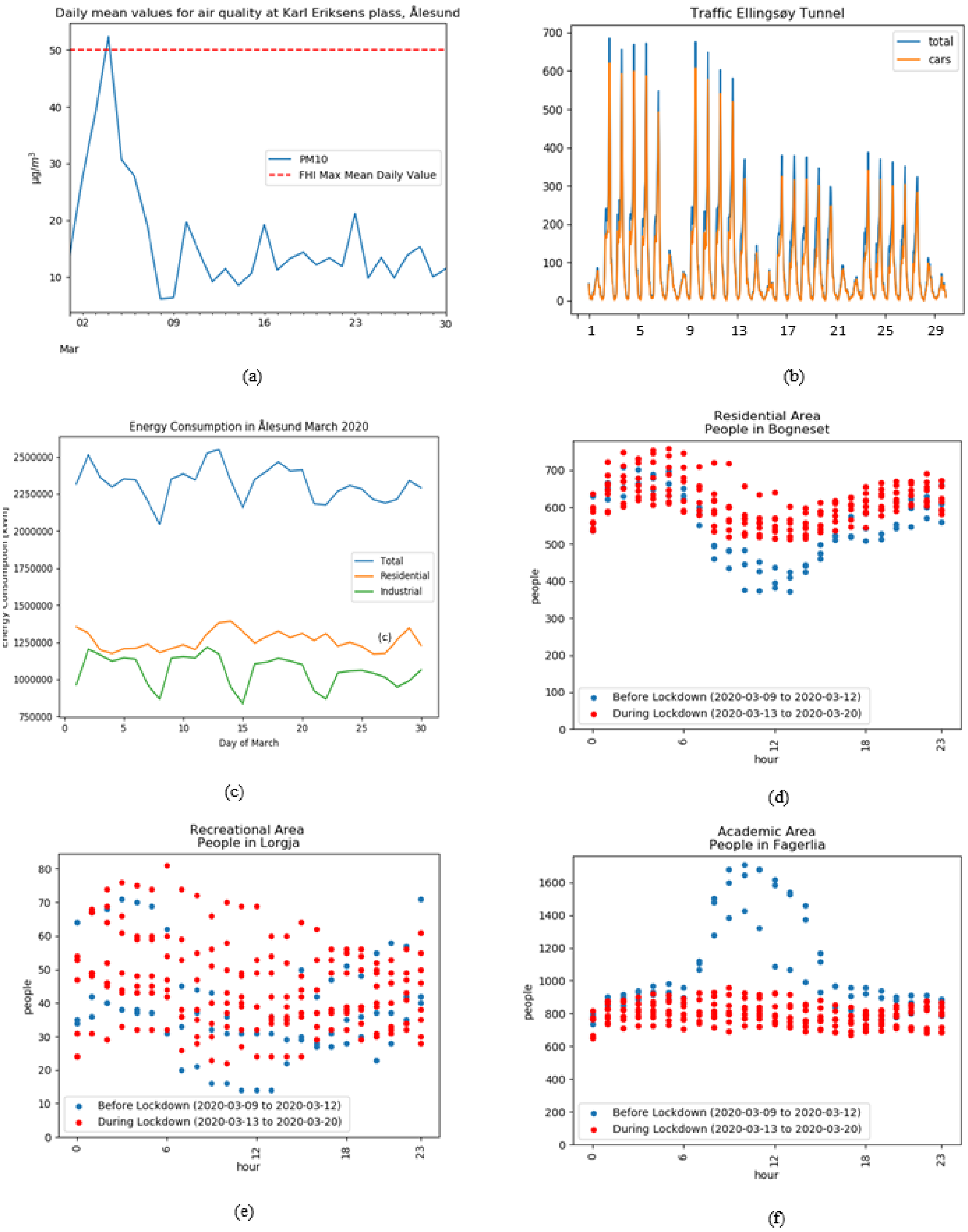


Figure 5: PM10 in Ålesund centre, March 2020, NILU API (a), Traffic in Ellingsøy Tunnel, March 2020, Energy Consumption on the City during March 2020 (b), People in basic geographical units, Courtesy of Telia (d, e, f)

The architecture presented in Figure 2 is not particular to the city of Ålesund and can easily be scaled up to bigger cities and agglomerations, or regions. Indeed, most of the required data (height map, energy, orthophoto, traffic, pollution) are from national/international agencies.

Aggregated data allow a qualitative interpretation of the patterns, but more fine-grain data would allow us to answer questions this paper could not. For example, we could not estimate quantitatively where the people live and work or which are their preferred recreational outdoor activities. Telia provides such origin-destination matrices, the data will be investigated in future studies. Likewise, the traffic data from SVV is too coarse as only the sum of both directions is provided on the Rest API, limiting the reach of the presented approach.

The Covid-19 pandemic has created natural experiments in cities that would have been impossible to implement under normal circumstances. We argue that conventional data-driven prediction algorithms could not possibly have predicted the outcomes for the city traffic without human intervention and modelling within the given assumptions and input parameters. This suggests the importance of understanding the measured data and its context when modelling.

The case study of air pollution showed the limitations of the inspecting the data in such a narrow time window and further investigations need to be done. For example, using time series from past years, predictive models could be compared with the real data and visualised in the GDT.

Conclusion

In this paper, we present a GDT smart city framework to contextualize the insights of big data in time and space and illustrate the benefits with the case studies related to the Covid-19 lockdown. GDTs do not replace other analytical tools which can also prototype data analytics and perform visualisation. However, they are great storytelling tools for discovering insights, communicating them, and drawing awareness to the related KPIs. Many national agencies and other data-providers deliver high quality sensing and prediction data, based on modern metrology, that can be freely used for the common good or monetized for commercial purposes. By setting the data in a 4D context, visualisation exploits the potential of the trove of data. While measured traffic dropped by 30% during lockdown in Ålesund compared to the weeks before lockdown, no direct impact on air quality was evident, in terms of PM10. Future studies should examine the drivers of this result. Finally, a discussion of the measurement uncertainty and interrogation about whether to visualize it are put forward.

Acknowledgments

This work was supported by Offshore Simulator Centre, the Norwegian Research Council (NRC) "Data-driven ship models for rapid virtual prototyping", industrial PhD under Grant Number 285949, and NRC “Smart Plan” Grant Number 310056. We thank the Municipality of Ålesund, Norway, the Norwegian Roads Authority, the Norwegian Meteorological Institute, the Norwegian Department for Civil Protection, and the Norwegian Institute for Air Pollution Research for their support on their APIs, and Telia and Mørenett for kindly sharing aggregated data for the period of interest. The visualisation was performed using AugmentCity. Dina Aspen from NTNU drew the Smart Plan image.

Bibliography

[1] D. Petri, “Big data, dataism and measurement,” *IEEE Instrum. Meas. Mag.*, vol. 23, no. 3, pp. 32–34, May 2020, doi: 10.1109/MIM.2020.9082796.

[2] V. Scotti, “Big data or big (privacy) problem?,” *IEEE Instrum. Meas. Mag.*, vol. 20, no. 5, pp. 23–26, Oct. 2017, doi: 10.1109/MIM.2017.8036692.

[3] S. Kolozali *et al.*, “Observing the pulse of a city: A Smart city framework for real-time discovery, federation, and aggregation of data streams,” *IEEE Internet Things J.*, vol.

6, no. 2, pp. 2651–2668, Apr. 2019, doi: 10.1109/JIOT.2018.2872606.

[4] C. Esposito, A. Castiglione, F. Frattini, M. Cinque, Y. Yang, and K. K. R. Choo, “On data sovereignty in cloud-based computation offloading for smart cities applications,” *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4521–4535, Jun. 2019, doi: 10.1109/JIOT.2018.2886410.

[5] Z. Li, R. Al Hassan, M. Shahidehpour, S. Bahramirad, and A. Khodaei, “A Hierarchical Framework for Intelligent Traffic Management in Smart Cities,” *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 691–701, Jan. 2019, doi: 10.1109/TSG.2017.2750542.

[6] F. G. Brundu *et al.*, “IoT software infrastructure for energy management and simulation in smart cities,” *IEEE Trans. Ind. Informatics*, vol. 13, no. 2, pp. 832–840, Apr. 2017, doi: 10.1109/TII.2016.2627479.

[7] M. Ward, G. Grinstein, and D. Keim, *Interactive data visualization: foundations, techniques, and applications, Second Edition (360 Degree Business)*. Boca Raton, FL: CRC Press, 2015.

[8] H. P. Halvorsen, O. A. Grytten, M. V. Svendsen, and S. Mylvaganam, “Environmental Monitoring with Focus on Emissions Using IoT Platform for Mobile Alert,” in *2018 28th EAEEIE Annual Conference, EAEEIE 2018*, 2018, doi: 10.1109/EAEEIE.2018.8534197.

[9] M. Ferro, V. Paciello, A. Pietrosanto, and P. Sommella, “A distributed measurement system for the estimation of air quality,” *IEEE Instrum. Meas. Mag.*, vol. 23, no. 5, pp. 51–56, Aug. 2020, doi: 10.1109/mim.2020.9153575.

[10] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva, “Visual exploration of big spatio-temporal urban data: A study of New York city taxi trips,” *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2149–2158, 2013, doi: 10.1109/TVCG.2013.226.