Smart City - smart data? Towards a holistic system of insight for data analytics in smart cities.

Marius Rohde Johannessen, Lasse Berntzen

University of South-Eastern Norway, Po Box 4, 3199 Borre, Norway.

Abstract

This paper discusses the potential insights of aggregated data and visualizations for decision-making. Our thesis is that the data is there but not always complete nor used as well as it should be. We argue for a "system of insight," where data from various sources are analyzed using machine learning and artificial intelligence, identifying deeper correlations between data from multiple sources. Such correlations would give decision-makers richer insights and a better understanding of relevant factors. To illustrate our case, we apply data from Norwegian municipalities to illustrate the potential of a holistic approach to data

Keywords

Big data, smart cities, machine learning, artificial intelligence

1. Introduction

eGovernment scholars have researched smart cities for almost a decade, and we have seen several exciting papers expanding the eGovernment knowledge base at conferences and in journals. There are many definitions of Smart Cities in the literature, ranging from general definitions to more comprehensive research and practical frameworks. For example, Giffinger [1] defines smart city as "a performing city built on the 'smart' combination of endowments and activities of self-decisive, independent, and aware citizens." Caragliu et al. [2] are a bit more concrete and mention the combination of human and social capital with technological development and the use of technology, while Dameri [3] defines smart cities in terms of elements (technology, citizens, land, and government), objectives, and boundaries and scope. Gil-Garcia et al. [4] present a framework showing how smart cities are made possible by combining technology and data, the physical environment, government, and society. Each of these dimensions is further operationalized. Most definitions seem to agree that smart city is a socio-technical phenomenon where we must examine technology in the context of objectives, physical conditions, human and organizational factors (etc.)

Technology and data are common parts of smart city frameworks and definitions. Many papers are looking into data collection, analysis, big data, and machine learning methods (See, for example [5]–[8]) and together these present us with a vast amount of research on effective data collection and analysis.

There is a need for more research on how to combine this with contextual factors, even though definitions of the smart city mainly emphasize "data in context" as essential. There are studies on data science and economic development, but not many are related to overall well-being or sustainability goals [9]. In the introduction to a special issue on smart cities, Meijer, Gil-Garcia, and Bolivar [10] stress the point that context, governance models, and a focus on the public value of smart city projects

are essential to succeed. Based on this, we propose that *there is a need for synthesis between research focusing on contextual issues and overall public value and research on the technical aspects of smart cities.*

Further, Smart cities often implicitly talk about big cities. When you google "smart city," you will find many results from cities such as Barcelona, Amsterdam, and London. Lopes and Oliveira [11] argue that smaller cities can also be smart. As there are a lot more small and medium-sized cities than large cities, and because Norway, where the data from this case is collected, is mainly made up of smaller cities, we also propose to

examine how smaller cities can apply data from various sources to gain deeper and richer insight, facilitating decision-making in areas related to smart cities.

This short paper examines how small and medium-sized smart cities can use data from various sources for insight and decision-making. We apply data from Norwegian municipalities as our case. Initial findings indicate that while large amounts of data are available, there is still a job regarding data governance and effective data utilization. We present preliminary findings based on interviews, participant observation, and documents. Our objective with this research is to create a framework for effective data science in smaller municipalities, addressing the smart city concept by linking data to context, governance structures, and public value. Here, we present our preliminary findings.

2. Towards a system of insight using interconnected data

The European strategy for data [12] points out that the data economy can be worth more than 800 billion euros in 2025 and defines the data economy as "value creation based on data as an important input factor in the production of goods and services, or when data is a driver for innovative solutions." The data strategy also emphasizes how data is essential for many public services and cities. Most importantly, the data strategy outlines a path toward a European single market for data, with common standards, regulations, and ethical guidelines, which could potentially be necessary for a more holistic approach toward data science in cities.

Jackson & Lockwood [13] explore how organizations use (big) data for decision-making and emphasize the need to think about data insights as a system comprising various data sources, workflows, governance, ethics, and other contextual and organizational factors. Inspired by this, we propose a "system of insight" based on both structured and unstructured data as a starting point for municipal data strategy (Figure 1).
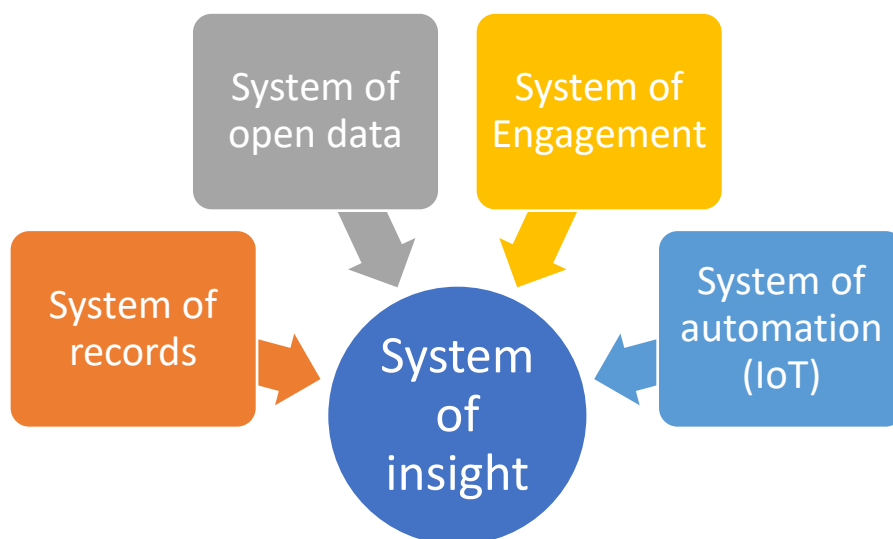

Figure 1: System of insight for smart cities. Inspired by Jackson & Lockwood (2018)

2.1.   System of records

Records refer to all the formal records recorded digitally by the government, such as taxes, health, vehicle registries, property and land records, environmental records, etc. In Norway, most cities use the Kostra database [14], a set of indicators for all municipalities' responsibilities, such as kindergartens, schools, and health. Kostra has economic data and data on selected indicators such as case handling time, school dropout rates, waste handling, and more[2]. These systems of records typically consist of structured data stored in databases and accessed using various applications.

2.2.   System of open data

Open data refers to government data that can be freely used, modified, and shared by anyone, for any purpose. Open data is typically structured, accessible in a machine-readable format, and accessed through various open data repositories [15]. These repositories consist of datasets with attached metadata and descriptions. In Norway, there are several open data repositories, which used to mean finding data was a challenge. With the introduction of the common data catalogue at data.norge.no, access to open data became easier, as the repository now links to most of the open data repositories in Norway. There are currently 1662 data sets and 192 APIs registered.

In 2021, the government white paper "data as a resource" was published, emphasizing the value of data and laying out a strategy for better use of public data sets. (Norwegian ministry of local government and modernization, 2021). Here, the government outlines four principles for Norwegian open data:
1.   Data should be open as standard and screened/restricted only when needed.
2.   Data should be available, searchable, linkable, complete, and easy to use.
3.   Data should be shared and used to facilitate value-creation for private/public sector and society.
4.   Data should be shared and used in such ways that fundamental rights and freedoms are respected, and Norwegian societal values are preserved (trust and ethics).

2.3.   System of engagement

Engagement in this context refers to citizen engagement on policy and social levels relevant to the city, such as letters to the editor in local newspapers and activity in geographically based social media groups. In the eGovernment context, engagement is roughly equivalent to eParticipation [16]. Engagement can also come through other channels, such as sports and voluntary organizations, but input from these would typically come through other channels. We can tap into these systems of engagement using various data mining techniques [17]. However, access is becoming more restricted with Facebook closing access after the Cambridge Analytica case and Twitter recently having announced limitations [18]. Most municipalities, politicians, and voluntary organizations in Norway are active social media users. A report from the municipal coordinating organization KS showed that Norwegian municipalities were active social media users [19]. In addition, local newspapers remain a source of information and discussion for many citizens, and social and traditional media supplement each other [20].

2.4.   System of automation

Automation refers to using the Internet of Things (IoT) devices, typically sensors, that collect data, connect, and exchange information via the Internet (Figure 2). According to Gartner, an estimated 20 billion devices are connected to the Internet [21]. IoT involves various devices, from air quality sensors to airplane jet engines. From a smart city perspective, devices with sensors and actuators are especially relevant, as these can measure different variables and make things happen based on data input. For example, air quality sensors can sense that air quality is low due to particulate matter, send a signal to

---

[2] Full list of Kostra indicators, see: https://www.ssb.no/en/statbank/list/kostrahoved

the traffic system and have digital speed signs lower the speed limit to decrease the amount of particulate matter being thrown around by spinning wheels, or have displays that inform citizens no one with a diesel engine is allowed within a specific area until the air clears.

In Norway, we have several examples of IoT applied in a smart city context. Examples include the NTNU-led project on zero-emission neighborhoods, relying on IoT for data on emissions, transport patterns, and other data relevant to cutting emissions [22]. Stenstavold et al. [23] conducted a study across several Norwegian municipalities. They found that sensors for air, water, waste, transport, buildings management, and fire safety were among the most common applications of IoT.
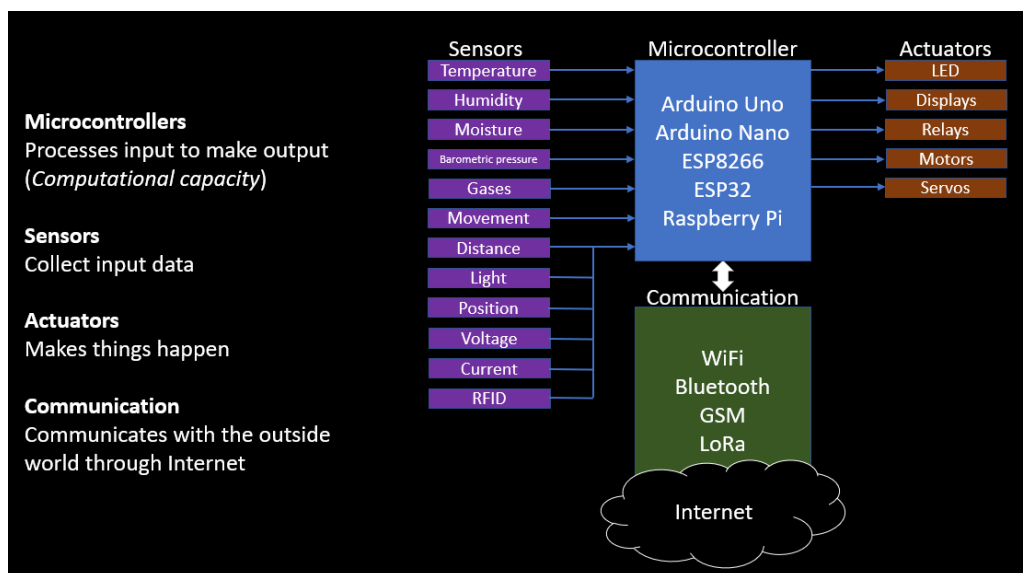

Figure 2: Illustration of IoT and how different devices connect to make things happen. Own illustration

2.5. Connect the dots - sourcing and analyzing data from multiple sources

We would argue that smart decision-making uses a systematic approach to data collection and applies logical decision-making techniques instead of using intuition, generalizing from experience, or trial and error. While there are many examples of open data, social media data, registry data, and IoT data being used in smart city contexts individually, i.e., Bilal et al. [7], there is less research into how we can create a system for combining these data sources and the potential value this could have for cities. For example, registry data from Kostra allows cities to compare how much they spend on healthcare or how schools are performing but provides little information about how or why. To answer that, we might need to supplement with other data sources such as social media and newspaper discussions on the quality and perceived satisfaction with municipal healthcare. Some systemic issues might need addressing if spending is high and satisfaction low. Open data could provide additional information. For example, 45 open data sets related to health, such as the location of institutions, could be combined with open data on environmental issues and air quality. Data from IoT sensors could also be of help. For example, tracking the movement of home care nurses and helping them find the optimal route between patients to save time. Tracking can also show the progress of removing snow after a snowfall or the location of garbage trucks when collecting garbage. The monitoring provides citizens with insights into the punctuality of service deliveries.

The above is just one example of combining different data sources to offer additional insight. The point is that we have potential access to massive amounts of data, but these are rarely used in combination. If we apply a combination of machine learning, statistical techniques, and artificial intelligence, a good analytical engine could help us discover novel insights, which could then be presented to decision-makers using data visualization dashboards (Figure 3). The benefit of machine

learning is that one or a few generic algorithms can be used to solve different problems. Clustering can help identify previously unknown relations, association rules algorithms can help us discover relations such as the example mentioned above, and data reduction techniques can be applied to large data sets to aggregate the information into more understandable chunks [5].
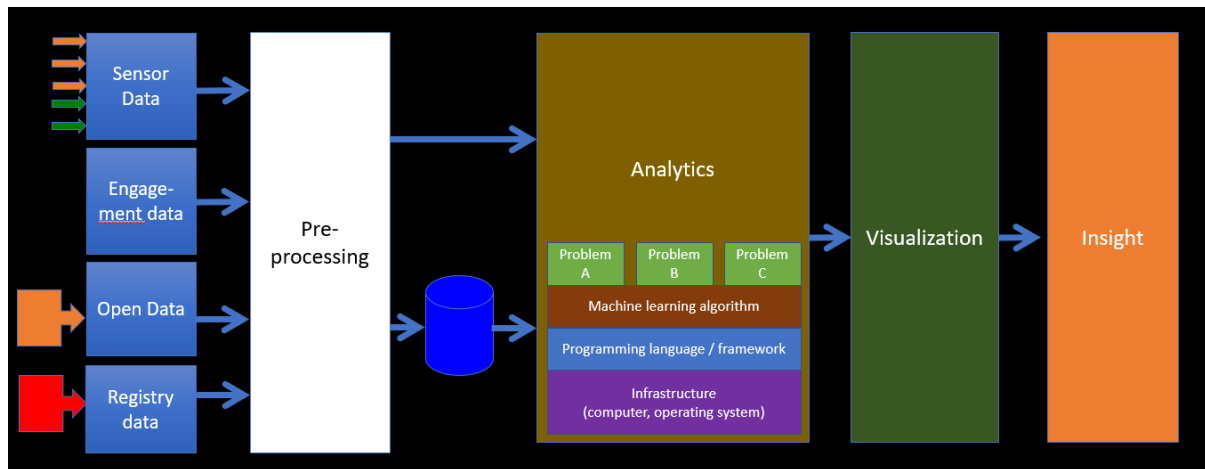

Figure 3: Conceptual model for the system of insight

Another aspect that could be added to the framework is assessment and evaluation. Patrão et al. [24] conducted a review of smart city evaluation frameworks and identified several frameworks which could be relevant to data-driven smart cities. They also summarized potential benefits for various stakeholder groups. Better and more detailed data would benefit both citizens, government, investors, and businesses.

There are many challenges and a lot of work remaining, which we will address in the conclusion. Still, first, we will present some examples from municipalities in our region in Norway in section 3.

3. Example case - Municipalities in Vestfold and Telemark County

In our region in the south of Norway, we have eleven cities, with a population ranging from 13.000 to 102.000. The region is interconnected with people living and working across different cities, as distances are relatively small. It takes a little over an hour to drive from south to north in the region. Many isolated examples of potential data sources could be combined to create a holistic system of insight. We have interviewed, observed, or participated in several of these examples.

Registry data: As mentioned above, all Norwegian municipalities rely on Kostra data when evaluating themselves or preparing the annual budget. The national property registry and tax data are other examples. There are also citizen surveys that provide feedback to the city regarding how satisfied citizens are with public services, although these are not conducted annually in all municipalities. Statistics Norway is responsible for Kostra and also has statistics for a lot of city-relevant indicators.

Open data: There are many open data sets relevant to cities, business, nature, environment, transport, property, families and children, health, culture, labor, geographic data, and education. Several of these can provide valuable supplemental information, perhaps especially data on water, flooding, and energy from the water resources and energy directorate, geographic data from the mapping authority, and environmental data from the environment agency.

Engagement data: Several mayors in the region's cities are active participants in social media, and all of them have local newspapers discussing politics and current affairs. There are also local Facebook groups where citizens meet and discuss everything from burglaries via dinner recipes to urban planning.

IoT data: We have several examples of IoT data in the region, and this is where we have begun our data collection and where we have drawn our initial findings and thoughts about the model presented in Figure 3.

There are several exciting initiatives in our cities, such as "Fjord city," a new district in the city of Drammen which can house 16.000 people when completed. Fjord city is planned and built with sustainability and the environment in mind, including IoT and sensing in traffic management and energy. The city of Horten is working on several IoT initiatives, the first being a project for monitoring water, stormwater, and sewage. Other examples on a more individual level include the "powerhouse" building, designed to produce more energy than it consumes and feed surplus energy back to the grid or directly to nearby buildings. The building has many sensors for optimizing energy use. There are also other examples in the region of solar panels combined with sensors, batteries, and neighboring businesses to optimize energy use. Our university is partnering up with the county council on an electric bike-sharing scheme, where GPS sensors could be used to track travel patterns. Here we could also include data from private micro-mobility scooter companies. For waste management, the "Magic factory" uses bio-waste from food and agriculture to produce biofuel used in regional public transport. The factory is an advanced and technology-heavy operation, working with several cities in the county and collaborating with schools in STEM education through its own learning center.

4. Discussion and conclusion - future research directions

In this paper, we have presented the various categories of data that could provide input for decision-makers in smart cities. We have also presented data sources, including IoT projects, used in our southern Norway region. Data is heavily applied in municipalities already, but few, if any, apply a holistic approach or include several data sources and machine learning for exploration and depth. Thus, we have proposed a model for a system of insight where a (big) data analytics lab could be set up to import data from registries, open data sets, engagement data, and IoT, apply various machine learning analyses, and present the results using data visualization techniques so that decision-makers have better access to richer data. The examples we present in section 3

Of course, our proposed model is just a starting point, and there is a need for a lot more research and development, and we would appreciate input regarding this at the conference. For now, we see the following avenues for future research:
- Stakeholder mapping and involvement. Who are the benefactors, contributors, and secondary users, and how should they be involved in design and use?
- Mapping relevant areas where richer data would benefit the city and mapping potential data sources for each area.
- Cost and project management. Setting up data labs can be expensive and consume a lot of resources. Machine learning and AI is energy intensive. Thus, research is needed into how we can lower energy consumption, sustainably use resources, and organize and manage such a project. For small smart cities, collaboration is probably needed. Relevant evaluation frameworks and evaluation of public value creation would likely be part of a socio-technical study of organizational factors.
- Creating data and visualizations is one thing, but having decision-makers actually use the results is another matter. The public sector already reports on an extensive range of indicators, and we often hear public officials moan about the reporting needs. Thus, research is needed into organizing and managing this and how we can better use the data. For example, strategies, standards, and processes for effectively using data are exciting avenues for further research.

In conclusion, while there is a massive amount of research on smart cities, data use, and evaluation, there is still a need for a holistic approach to systematic data collection, analysis, and use in smart cities.

5. Acknowledgements

6. References

[1] R. Giffinger, C. Fertner, H. Kramar, and R. Kalasek, "Smart cities - Ranking of European medium-sized cities," 2007. Accessed: Jun. 27, 2023. [Online]. Available: https://www.researchgate.net/publication/261367640_Smart_cities_-_Ranking_of_European_medium-sized_cities

[2] A. Caragliu, C. Del Bo, and P. Nijkamp, "Smart Cities in Europe," *Journal of Urban Technology*, vol. 18, no. 2, pp. 65–82, 2011, [Online]. Available: http://ezproxy.hbv.no:2158/ContentServer.asp?T=P&P=AN&K=64133864&S=R&D=aph&EbscoContent=dGJyMNLe80SeqK44zOX0OLCmr0yeprdSsa64SLeWxWXS&ContentCustomer=dGJyMPGrtkiwqLdLuePfgeyx44Dt6fIA

[3] R. Dameri and R. P. Dameri, "Searching for Smart City definition: a comprehensive proposal Searching for Smart City definition: a comprehensive proposal TYPE (METHOD/APPROACH) Council for Innovative Research," *Peer Review Research Publishing System Journal: International Journal of Computers & Technology*, vol. 11, no. 5, 2013, doi: 10.24297/ijct.v11i5.1142.

[4] J. R. gil-Garcia, T. A. Pardo, and T. Nam, "What makes a city smart? Identifying core components and proposing an integrative and comprehensive conceptualization," *Information Polity*, vol. 20, no. 1, pp. 61–87, 2015.

[5] S. El Mendili, Y. El Bouzekri El Idrissi, and N. Hmina, "Benchmarking study on smart city data analytics," *Colloquium in Information Science and Technology, CIST*, vol. 0, pp. 841–846, Jul. 2016, doi: 10.1109/CIST.2016.7805004.

[6] A. M. S. Osman, "A novel big data analytics framework for smart cities," *Future Generation Computer Systems*, vol. 91, pp. 620–633, Feb. 2019, doi: 10.1016/J.FUTURE.2018.06.046.

[7] M. Bilal *et al.*, "Smart Cities Data: Framework, Applications, and Challenges," *Handbook of Smart Cities*, pp. 1–29, 2020, doi: 10.1007/978-3-030-15145-4_6-1.

[8] I. H. Sarker, "Smart City Data Science: Towards data-driven smart cities with open research issues," *Internet of Things*, vol. 19, p. 100528, Aug. 2022, doi: 10.1016/J.IOT.2022.100528.

[9] S. E. Bibri, "On the sustainability of smart and smarter cities in the era of big data: an interdisciplinary and transdisciplinary literature review," *J Big Data*, vol. 6, no. 1, pp. 1–64, Dec. 2019, doi: 10.1186/S40537-019-0182-7/TABLES/3.

[10] A. J. Meijer, J. R. Gil-Garcia, and M. P. R. Bolívar, "Smart City Research," *http://dx.doi.org/10.1177/0894439315618890*, vol. 34, no. 6, pp. 647–656, Dec. 2015, doi: 10.1177/0894439315618890.

[11] I. M. Lopes and P. Oliveira, "Can a small city be considered a smart city?," *Procedia Comput Sci*, vol. 121, pp. 617–624, Jan. 2017, doi: 10.1016/J.PROCS.2017.11.081.

[12] European commission, "European data strategy - Making the EU a role model for a society empowered by data ," 2017. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en (accessed Jun. 16, 2023).

[13] T. W. Jackson and S. Lockwood, *Business Analytics: A Contemporary Approach*, vol. 275. London: Springer nature Publishing, 2018. doi: 10.1016/J.INS.2014.01.015.

[14] SSB, "KOSTRA," 2023. https://www.ssb.no/offentlig-sektor/kostra/statistikk/kostra-kommune-stat-rapportering (accessed Jun. 27, 2023).

[15] Y. Charalabidis, A. Zuiderwijk, C. Alexopoulos, M. Janssen, T. Lampoltshammer, and E. Ferro, "The World of Open Data," vol. 28, 2018, doi: 10.1007/978-3-319-90850-2.

[16] Ø. Sæbø, J. Rose, and L. Skiftenes Flak, "The shape of eParticipation: Characterizing an emerging research area," *Gov Inf Q*, vol. 25, no. 3, pp. 400–428, 2008, [Online]. Available: http://www.sciencedirect.com/science/article/B6W4G-4P7FCVJ-1/2/f6868c8dc660d4308bbbd06abfcab4ad

[17]   O. Belkahla Driss, S. Mellouli, and Z. Trabelsi, "From citizens to government policy-makers: Social media data analysis," *Gov Inf Q*, vol. 36, no. 3, pp. 560–570, Jul. 2019, doi: 10.1016/J.GIQ.2019.05.002.

[18]   K. Kupferschmidt, "Twitter's threat to curtail free data access angers scientists | Enhanced Reader," *Science*, vol. 379, no. 6633. New York, 2023.

[19]   O. E. Johnsen, E. C. Jørgensen, and R. E. Aaneland, "HVORDAN LYKKES KOMMUNESEKTOREN I SOSIALE MEDIER?".

[20]   M. P. Lie, "Local newspapers, Facebook and local civic engagement: A study of media use in two Norwegian communities," *Nordicom Review*, vol. 39, no. 2, pp. 49–62, 2018, doi: 10.2478/NOR-2018-0011.

[21]   M. Hung, "Leading the IoT: Gartner insights on how to lead in a connected world," 2017.

[22]   T. K. Soltvedt, A. Sinaeepourfard, and D. Ahlers, "A Cost Model for Data Discovery in Large-Scale IoT Networks of Smart Cities," *Proceedings - IEEE International Conference on Mobile Data Management*, vol. 2020-June, pp. 348–353, Jun. 2020, doi: 10.1109/MDM48529.2020.00076.

[23]   M. Stenstadvold, "Smarte byer og kommuner i Norge-en kartlegging KOMMUNAL-OG MODERNISERINGSDEPARTEMENTET", Accessed: Jun. 27, 2023. [Online]. Available: www.stock.adobe.com

[24]   C. Patrão, P. Moura, and A. T. de Almeida, "Review of Smart City Assessment Tools," *Smart Cities 2020, Vol. 3, Pages 1117-1132*, vol. 3, no. 4, pp. 1117–1132, Sep. 2020, doi: 10.3390/SMARTCITIES3040055.