

## Secondary Publication



Nicklas, Daniela; Ackermann, Leonie; Das, Debasree; u. a.

### Challenges for Federated Crowd Management in Smart Cities

Date of secondary publication: 16.01.2026

Version of Record (Published Version), Conferenceobject

Persistent identifier: urn:nbn:de:bvb:473-irb-112571x

#### Primary publication

Nicklas, Daniela; Ackermann, Leonie; Das, Debasree; u. a. (2026): Challenges for Federated Crowd Management in Smart Cities, in: Keiichi Yasumoto, Idit Keidar, Hirozumi Yamaguchi, u. a. (Ed.), ICDCN Companion '26: Companion Proceedings of the 27th International Conference on Distributed Computing and Networking, New York: Association for Computing Machinery, pp. 126–131, doi: 10.1145/3737611.3776621.

#### Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:

<https://creativecommons.org/licenses/by/4.0/legalcode>

# Challenges for Federated Crowd Management in Smart Cities

Daniela Nicklas  
Universitaet Bamberg  
Bamberg, Germany  
daniela.nicklas@uni-bamberg.de

Leonie Ackermann  
Universitaet Bamberg  
Bamberg, Germany  
leonie.ackermann@uni-bamberg.de

Debasree Das  
Universitaet Bamberg  
Bamberg, Germany  
debasree.das@uni-bamberg.de

Tatsuya Amano  
The University of Osaka  
Osaka, Japan  
t-amano@ist.osaka-u.ac.jp

Hamada Rizk  
The University of Osaka  
Osaka, Japan  
hamada\_rizk@ist.osaka-u.ac.jp

Hirozumi Yamaguchi  
The University of Osaka  
Osaka, Japan  
h-yamagu@ist.osaka-u.ac.jp

## Abstract

This paper addresses the issue of crowd management in smart cities, where effectively handling large groups is essential for ensuring safety and a positive experience for all. Although smart cities strive to use data for better decision-making, simply measuring crowds is not enough; cities must analyze and utilize this data intelligently. In this paper, we leverage the various experiences of two research groups in this area and present a vision that allows cities to easily establish crowd management platforms by transferring knowledge and insights from similar use cases in other cities.

## CCS Concepts

- **Computer systems organization** → **Data flow architectures**; • **Information systems** → *Combination, fusion and federated search*;
- **Hardware** → *Sensor applications and deployments*.

## Keywords

smart city;crowd monitoring;federated learning

### ACM Reference Format:

Daniela Nicklas, Leonie Ackermann, Debasree Das, Tatsuya Amano, Hamada Rizk, and Hirozumi Yamaguchi. 2026. Challenges for Federated Crowd Management in Smart Cities. In *Companion volume of the 27th International Conference on Distributed Computing and Networking (ICDCN Companion 2026)*, January 06–09, 2026, Nara, Japan. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3737611.3776621>

## 1 Introduction

The vision of smart cities and communities is to harness data for better decision-making, moving from passive monitoring to actively shaping efficient, resilient, and responsive urban environments. While there is no universally accepted definition of a smart city [7], the concept encompasses social, economic, and ecological dimensions. Crowd management illustrates this potential, as overcrowding can lead to negative experiences and safety risks. Data-driven solutions are needed to support applications such as managing hot spots, ensuring safety, balancing visitor flows, and aligning mobility services with demand. Cities face challenges including costs,

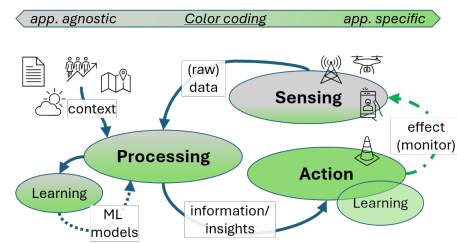


Figure 1: Smart Systems Cycle

privacy concerns, and potential misuse. While some larger cities may have established smart systems, many are just starting out, lacking experience and funding. Additionally, machine learning needs historical data to train high-quality models, leading to long ramp-up time for new installations before benefits are realized.

Figure 1 illustrates the general cycle of smart systems. It is adapted from the well-known Sense-Think-Act paradigm from robotics, which is also useful to discuss pervasive systems [3]. A smart system needs to sense the current situation (e.g., counting the number of people at a specific location). This raw data is combined with contextual information (such as maps, event schedules, or historical records) and then processed—for instance, to train machine learning models, predict critical situations, or conduct statistical analyses. The resulting insights support decision-making, which may lead to automated actions (e.g., adjusting traffic light cycles) or human-in-the-loop interventions (e.g., deploying additional security staff). The effects of these actions need then to be monitored, which can be done by sensing again. While sensing and some processing can be application-agnostic, the actions always depend on the application. In this paper, we discuss how a distributed multi-sensory approach could address these challenges. Distribution is key, both spatially (within the city, between cities) and organizationally (different organizations like event organizers, city councils, or mobility managers). Based on experiences from real-world applications in research groups, our main contribution is a vision on how a combination of different sensor technologies and federated transfer learning can reduce the cost and time needed for smart cities to establish efficient crowd management systems. Additionally we propose a conceptual architecture and outline prior work contributing to achieving our vision. The rest of the paper is organized as follows: We will first detail the challenges for distributed crowd management and then outline the vision in Chapter 3. Chapter 4



This work is licensed under a Creative Commons Attribution 4.0 International License. *ICDCN Companion 2026, Nara, Japan*  
© 2026 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-1969-1/26/01  
<https://doi.org/10.1145/3737611.3776621>

reports on previous work. We discuss the feasibility of our approach in Chapter 5 before we conclude the paper and outline future work.

## 2 Challenges

In this chapter, we first highlight the primary challenges that must be addressed across the three key stages of a smart system: Sensing, Processing, and Action, as illustrated in Figure 1.

**Sensing.** Various sensing technologies are available to assess crowd density in a given location, including vision-based systems (cameras), pressure sensors, LiDAR, WiFi signals, and mechanical people counters. Selecting the most suitable technology is a complex task, as it must address several challenges. Although many sensors can serve multiple applications, the final choice may be influenced by the requirements of the most demanding application:

**Cost:** The sensing solution must remain within budget constraints, considering both installation and maintenance expenses. **Data Quality:** The technology should deliver data that meets the necessary quality standards for the application, covering value accuracy (the precision of the count), spatial accuracy (the precision of location sensing), spatial completeness or coverage (the extent to which the target area is monitored and the presence of blind spots), and timeliness (the frequency of updates and associated latency). **Privacy:** It is essential to assess how well the privacy of individuals is protected. Specifically, one should consider whether it is feasible to re-identify individuals based on the collected data, especially in non-crowded scenarios. **(Re)liability:** The reliability of the system is paramount, as is the question of accountability for sensing errors. This is particularly critical in safety-related applications, e.g. to restrict the maximum number of people. **Technical Sovereignty:** Understanding who has the authority to operate or shut down the system is crucial, especially in a landscape where international technology companies may be subject to regulations from their home countries that do not align with the interests of their users.

**Processing.** In the processing step, sensor data is utilized to generate the information required for specific applications. While certain aspects of processing, such as data cleaning and anonymization, are general and applicable across various contexts, the majority of the processing is typically tailored to meet the needs of individual applications. Here, we need to address the following challenges:

**Data Heterogeneity:** Different sensor systems deliver information in various formats. For example, LiDAR provides point clouds, WiFi sensing captures data from WiFi probes received in a specific area, light barriers measure the number of obstacles (such as people) crossing a line, and data collection applications deliver GPS location trajectories. While utilizing multiple data sources is recommended to enhance data quality, this approach necessitates addressing the resulting heterogeneity. **Latency:** Certain applications require rapid responses to changing situations, e.g., reacting to overcrowding to avoid panic. In these cases, not only must the sensing be real-time, but the processing also needs to deliver insights with low latency. **ML model training:** Since many steps in sensor data processing are now performed using AI, it is essential to train ML models that are reliable and efficient. Such models should be effective for the intended task, avoid overfitting, require neither excessive computational resources nor energy, and remain robust

against concept drift when deployed in real-world environments. Achieving this can be particularly challenging in newly installed sensor systems, where only limited historical data are available and seasonal variations are not yet fully captured. This limitation directly motivates the next challenge, **ML model management:** In real-world environments, conditions evolve over time. Consequently, ML models that were initially trained and deployed may no longer remain optimal for operation. It is therefore necessary to continuously monitor their predictive performance and, when required, adapt, replace, or retrain the models to ensure sustained accuracy and reliability. **Distributed Processing:** We need to allocate the processing to a physical location: At the edge (e.g., directly at the sensor or in its vicinity), at the backend (such as a server within the city’s infrastructure or in the cloud), or at intermediate nodes (commonly referred to as fog computing)? Processing may also occur near the actuator itself (e.g., close to an automated door or dashboard). Addressing this challenge requires careful system design that considers not only optimal resource allocation, but also non-functional requirements such as privacy, network latency, or resilience.

**Action.** When the processing of sensor data yields relevant information, an *action* refers to a measure taken to manage the crowd. Depending on the application, such actions may be automated (e.g., adjusting the schedule of a traffic light or closing an entry) or human-controlled (e.g., deploying additional security personnel to a specific area or adapting the bus schedule). The design of these actions entails multiple challenges that must be addressed:

**Human autonomy:** Actions in crowd management often aim to influence human behavior, for example by enforcing certain mobility, providing direct communication (e.g., public messaging), or using indirect cues (commonly referred to as nudging). Nevertheless, individuals retain free will and may not respond as intended. **Information delivery:** Even if individuals are willing to follow the recommendations or instructions of the crowd management system, they must first receive the message. The challenge lies in selecting an appropriate communication channel: not everyone owns a smartphone or is willing to install an application, public displays may be restricted in historic sites, and audio announcements can only be delivered efficiently in a limited number of languages. **Information uncertainty:** The information produced by data processing may contain inherent uncertainty, which can arise from low sensor quality, limited prediction accuracy, or a combination of both. Nevertheless, the crowd management application must respond to such information, giving rise to the challenge of decision-making under uncertain conditions. **Conflicts of interest:** In a city, multiple stakeholders or organizations may have an interest in crowd management, either to enhance their services or to comply with legal requirements. Examples include the city administration, the tourism board, public transport agencies, the economic development department, or law enforcement. However, the objectives of these stakeholders may differ, which in turn affects the actions that should be taken when overcrowding is detected: while public safety can be improved through better crowd distribution, local businesses at popular locations may benefit from higher foot traffic. **Policy optimization:** To effectively manage crowd flow based on congestion levels, it is essential to design optimal control policies that

integrate both macroscopic and microscopic approaches. For instance, at stations or event venues, macroscopic control can involve guiding the overall flow of people to reduce congestion. In contrast, microscopic control can be achieved by providing individuals with incentives via smartphones to adjust their behavior voluntarily. By combining large-scale guidance with personalized interventions, we need to optimize the overall control strategy to maximize both collective efficiency and individual satisfaction.

### 3 Vision

In our vision for federated crowd management, some smart cities act as leaders. They might have leveraged public funding from research projects or urban innovation programs, or have local stakeholders who invested in crowd management applications. Since they are typically big cities, they might also have crowd management deployed in several locations within their city limits. Other smart cities or local stakeholders might want to follow. By joining the federated platform we propose, they can leverage the efforts and experiences from the leading cities and use pretrained models for their local processing, speeding up the time from installation to application significantly. In Figure 2, we depicted a conceptual architecture for our vision. It consists of four layers (L0 to L3) and shows three cities (a, b, and c) that deployed crowd management applications (a1 and a2) to different locations.

**Layer L0** is the layer that is closest to the physical world. Different sensing systems (e.g., LiDAR, WiFi-trackers, or GPS trajectories from mobile tracking applications) can be used to measure the presence of people. In our example, city a is a leader city. It has three locations covered with sensor installations (a.11 to a.13), with two of them spatially overlapping. Two other cities (b and c) have only one location covered yet.

Sensor-specific processing occurs at **layer L1**. We propose that it be executed as close to the data source as possible to reduce network latency and potentially enhance security and privacy. At this layer, processing includes automated data quality control, anonymization, and aggregation. This is also where data is unified into a common crowd data model, enabling it to be fused and processed collectively (e.g., using named locations or grid cells with crowd density levels). If such integration is not yet feasible (as in the case of city B), it should be implemented as a separate processing unit within the city platform, facilitating easier testing and verification, and ensuring consistent pre-processing and anonymization across all applications.

**Layer L2** consists city-specific processing. Depending on the applications or actions the city wants to deploy for crowd management, several processing steps can happen. They typically involve the training of ML models that can be used to predict relevant situations, which require action. In our figure, we show two different applications a1 and a2 (e.g.: safety for large events and data-driven on-demand mobility) that are deployed to different areas in the sample cities a, b, and c. The actions have an effect back to the real world which should be monitored again by the sensing layer L0.

Without **layer L3**, our architecture would reflect the state of the art in current smart cities. However, since many smart cities run similar applications—be it because they learned from each other, or because they try to target the same legal requirements—we propose

another layer that contains cross-city processing. The goal of this layer is to transfer knowledge and trained models from one city to another, or between locations within the same city. This process can be achieved through transfer learning or domain adaptation, where a machine learning (ML) model is initially trained using data from one location and subsequently adapted using historical data from another location. However, this approach not only necessitates access to historical data from the new target location—data that may not be readily available, but it can also be difficult to implement in a federated manner. Therefore, we propose an alternative approach that leverages federated domain generalization. In this approach, cities train their local models on dataset L2, which are then aggregated to create a global, region-neutral model on level L3. This technique is analogous to the methodology outlined in Ozeki et al. (2023) [13]. The resulting model can then be deployed in previously unseen target locations where historical data is lacking. In our example, this could be city c, which might have installed just recently their sensor system to monitor and manage the crowds at a big event (application 2), but has no historical data from similar events yet. It could ask city b whether it can join the consortium and thus use the region-agnostic model.

### 4 Previous Work

In this chapter, we shortly outline previous work from the authors or other groups that could contribute to the vision depicted in Chapter 3.

*Layer L0: Sensing.* Regarding crowd monitoring, various technologies can be employed. Camera-based solutions [17, 18, 26] offer a promising alternative to manual estimation. However, they raise privacy concerns and are also highly energy-intensive. Non-image-based localization solutions typically require active participation from individuals. In this case, either a smartphone application [16] must be installed, or self-built devices equipped with RFID tags [11] or Bluetooth technology [23] must be voluntarily carried.

In [21], we address the concern of individual identification through RGB video footage for pedestrian tracking with an alternative of LiDAR-based point-cloud data. This is particularly interesting in cases where fast-moving people are often obstructed by luggage or strollers during crowded scenarios. Another alternative to predict the direction of crowd movement is to process the halting or changing stations from intermodal anonymized trajectory data collected at city scale using a smartphone app funded by the Explanym project for Bamberg city. Although GPS data is vulnerable, specifically revealing sensitive information about users, our previous work [5] shows that if trajectory data can be anonymized considering the mode of transportation, leveraging differential privacy, it preserves important information such as detecting mode of transport or counting people within a predefined radius for intermodal data. Another data source could be crowdsensing approaches using tailored sensors like the OpenBikeSensor<sup>1</sup> and use donated information about overtaking distances (safe/unsafe) and bicycle trajectories to detect busy street localization.

In another study [1], we deployed a set of WiFi sensors (capturing WiFi probe requests) in tourist-relevant locations. Since many

<sup>1</sup><https://www.openbikesensor.org/docs/classic/> (Accessed Online: 2025-10-04, in German)

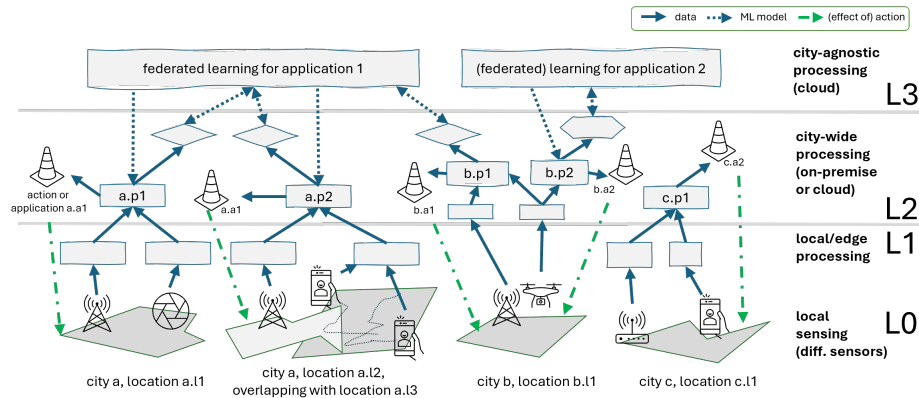


Figure 2: Conceptual Architecture and Sample Scenario

people carry at least one mobile device that can send out such signals, these devices can be used to detect crowding situations. Although it is known that these sensors have significant issues with data quality [15], they represent a cost-effective solution with low installation requirements. In contrast to many other crowd-sensing approaches, they do not require line-of-sight to the crowd. This is particularly useful for monitoring crowds in heritage areas, where any visible building modifications require formal approval.

*Layer L1: Sensor-specific processing.* In [2], we present a fully anonymized approach for processing WiFi sensor data, enabling edge (L1) processing. WiFi probe messages can be used to re-identify devices, even with MAC randomization in newer devices, which poses significant privacy risks by enabling tracking without consent from individuals in crowds. Therefore, our method not only hashes device identifiers (MAC addresses) but also deletes information from static devices and removes low-frequency periods that heighten re-identification risks. For instance, if a person is the only one walking through the city at night, their movements could be uniquely traced by a monitoring system. Our low-cost, non-intrusive sensors effectively count people in crowds while maintaining privacy. This system is currently implemented in Bamberg’s old town as part of the CrowdAnym project, providing data-driven visitor guidance. Experimental results indicate that our model performs comparably to laser scanners under moderate visitor density.

*Layer L2: City-wide processing and applications.* The research team at The University of Osaka has been developing a LiDAR-based crowd tracking system deployed across the campus [12] along with an AI model for congestion prediction. This system can capture the location, velocity, and trajectory of individuals within buildings at a fine spatial resolution, enabling detailed statistical analysis and spatiotemporal modeling of crowd density and movement. The team is also conducting a demonstration project on crowd management in a public park in Wakayama City [20]. The project aims to control visitors’ circulation behavior and to explore what types of content provision can increase visitors’ dwell time through cloud-based management. To this end, five LiDAR sensors have been installed in the park, and the data are integrated with privacy-preserving, temporally sparse GPS data collected from an Android smartphone

application to reconstruct human flow dynamics. Meanwhile, a simulation-based approach has been proposed to manage post-game congestion in a baseball stadium, aiming to alleviate the rush of spectators leaving the venue after games through AI-assisted optimization of crowd movement [27]. To address the computational cost of simulation-based methods, a surrogate model technique using neural networks has been developed for rapid crowd control optimization, demonstrated with Koshien Stadium data [19].

Although these approaches differ in purpose and scale, they share a common goal of crowd management. Because the sensing environments and system scales vary greatly across cities and applications, developing technologies that enable mutual utilization and integration of knowledge and AI models across different cities and applications represent an important research challenge.

*L3: Cross-city processing and federated approaches.* Cross-city transfer has become an important strategy to address data scarcity in target cities. One early and influential work is RegionTrans [22], which aligns regions across cities through an inter-city matching function and transfers spatio-temporal representations using deep models such as ConvLSTM with region-level layers, thereby improving prediction in data-scarce settings. Later, CrossTRes [8] introduced adaptive re-weighting of source city regions to reduce negative transfer. More recent efforts, such as Cross-IDR and multi-scale adaptation frameworks, focus on dynamic alignment and incremental distribution rectification to better capture shifting spatial dependencies [6, 24]. In parallel, federated and privacy-aware learning has emerged as a complementary paradigm for spatio-temporal prediction. Federated learning (FL) has been applied to traffic, mobility, and crowd flow forecasting, but challenges remain in handling heterogeneous data, communication costs, and the need for personalization [4]. To address these issues, **federated meta-learning** combines FL with meta-learning to adapt to non-IID spatio-temporal data by constructing global pattern graphs while supporting local customization [10]. Other works leverage graph neural networks in distributed settings; for instance, FLAGCN introduces asynchronous graph convolutional updates, while semi-decentralized STGNN training partitions computation across edge cloudlets to balance scalability, accuracy, and privacy [9, 25].

Building on these directions, our contributions tackle two remaining gaps: First, Ozeki et al. [13] proposed a region-neutral forecasting framework that eliminates the need for re-training when deployed in unseen cities. By disentangling region-specific and region-agnostic features via a variational autoencoder and embedding them within a graph neural backbone, this “one-model-fits-all” design achieves robust generalization across diverse urban contexts. Second, Ozeki et al. [14] introduced CC-Net, a federated architecture that embeds privacy protection at its core. Unlike conventional FL, CC-Net clusters clients by model similarity, applies local contrastive learning to mitigate imbalance, and is resilient to membership inference attacks. It has been verified that CC-Net not only preserves privacy but also improves accuracy beyond existing FL baselines.

## 5 Discussion

As we can see from the previous work, our vision presented in Chapter 3 appears feasible, as many of the challenges have already been addressed at least partially. However, there remains a gap that needs to be addressed through further research and engineering to fully enable effective and efficient crowd management in smart cities. In the following, we highlight the challenges in data representation, data quality, and city data platforms.

*Data representation.* Crowd sensing data can be represented in many ways, and the choice of sensing technology or L1 preprocessing pipeline often leads to variations in both feature sets and underlying distributions. For instance, one city may rely on WiFi sensing, while another uses LiDAR scans. Even within the same modality, differences in sensor density, placement, and calibration can introduce significant distributional shifts. Such heterogeneity hinders the effectiveness of federated learning (FL), as local models trained on non-aligned features or imbalanced data may contribute poorly to global aggregation, leading to degraded performance and biased outcomes across cities. FL paradigms provide different ways to cope with these challenges. Horizontal FL is suitable when cities collect similar features but from different populations, while vertical FL enables collaboration when overlapping populations are described by distinct features. In more complex and realistic scenarios, however, both the feature spaces and user populations may differ, making knowledge sharing far more challenging as heterogeneity arises across all dimensions. To address this challenge, we propose to normalize the crowd monitoring data so that it becomes comparable between locations and different cities on three dimensions: Crowdedness, Location and Time.

For measuring *crowdedness*, a straightforward representation would be the absolute number of people. However, some sensing technologies do not capture all individuals but only a subset (e.g., those carrying a WiFi device that has sent a probe, or citizens donating a GPS trajectory in a bicycle traffic collection campaign). Furthermore, the perceived crowdedness depends not only on the number of people present but also on the characteristics of the location (narrow, confined spaces vs. open areas). Therefore, we propose developing a categorical metric for “crowdedness” (e.g., low, medium, crowded, over-crowded) that should be standardized across cities. By doing this, we can homogenize information from

different sensor systems into a comparable representation. Regarding the *location*, one of these two representations are suited for the application: (i) Semantic spatial model with named locations (POIs): The relevant locations are represented as nodes in a spatial graph, crowd information is associated with these nodes. The graph captures distances and connections between locations. (ii) Normalized grid cells: Space is represented in grid cells (e.g., 100x100 meters), crowd information is associated with grid cells. However, it should be the same model within the same federation. In terms of *time*, the size of the smallest useful time window should be defined (e.g., 10 minutes). If data is needed in larger aggregations, these can be easily computed as long as they can be represented as multiples of the smallest time window.

*Data quality.* One could argue that data heterogeneity must not be a problem if we only rely on one sensor technology. However, each choice in this context may entail significant drawbacks, as you can see in Table 1 where we rated the criteria from Section 2 for some of the popular data sources for crowd management. Camera-based systems are cost-effective but offer high counting accuracy only in good lighting and have medium to high liability due to video evidence. However, they raise privacy concerns and affect public behavior which limits the deployment in public places. LiDAR systems, while more expensive, provide better counting accuracy regardless of lighting and improved privacy since they collect point clouds. WiFi sensing is low-cost with medium to high coverage but has lower counting accuracy and liability. Cellular Data depends on mobile phone signals from cellular providers, which can be costly to acquire and potentially lead to distorted data in overcrowded situations. This reliance also limits technical sovereignty, as crowd management effectiveness depends on the provider’s data availability. As we can see, no single technology provides crowd data of sufficient quality. Therefore, we conclude that reliable and cost-effective crowd management will inevitably require a combination of different sensing technologies, which can be used to compensate or calibrate the various signals.

*City Data Platforms.* Even if the challenge of data quality can be addressed through intelligent combinations of technologies to produce homogenized crowd monitoring data, a data platform is still required to manage this data both within individual cities and across multiple cities. Developing a city data platform as an open-source platform holds significant value from both technological and social perspectives. By making the underlying architecture and interfaces publicly available, interoperability across different cities can be ensured, preventing vendor lock-in and enabling cost-effective deployment even in smaller cities. Open sourcing also enhances transparency and accountability, as anyone can inspect how urban data is collected, processed, and utilized. It not only reduces development and maintenance costs but also promotes sustainable urban management. Finally, open-source platform initiatives contribute to global standardization and research advancement by aligning with international frameworks such as FIWARE and ISO/IEC Smart City standards. Another example is Civitas Core<sup>2</sup>, a collaboration

<sup>2</sup><https://www.civitasconnect.digital/civitas-core/> (Accessed Online: 2025-10-03, in German)

**Table 1: Comparison of sample sensing technologies based on key criteria**

Criteria	Camera	LiDAR	WiFi Sensing	Cellular Data
Cost	Low	Medium–High	Low	Medium–High
DQ: Counting Accuracy	Medium	High	Low	Medium
DQ: Spatial Coverage	Medium	Medium	Medium–High	High
DQ: Timeliness	High	High	Medium–High	Low
Liability	Medium–High	Medium	Low	Low
(Perceived) Privacy	Low	High	Medium–High	Medium–High
Technical Sovereignty	High	High	Medium–High	Low

of multiple German cities. We aim to integrate our concepts into such data platforms.

## 6 Conclusion

In this paper, we discussed the challenges of federated crowd management and proposed a federated learning–based framework concept for smart cities that integrates the entire cycle from sensing and processing to action. By enabling knowledge and model sharing across cities, the proposed approach has the potential to reduce deployment costs and shorten the training period for new implementations. As future work, we plan to conduct real-world demonstrations in cities in Japan and Germany to further enhance the framework’s capabilities in integrating heterogeneous sensing data and ensuring privacy-preserving learning, ultimately aiming to incorporate these concepts into open-source urban data platforms.

## References

- [1] Leonie Ackermann, Christoph Baum, Syed Ibrahim Khalil, Aleksandr Litvin, and Daniela Nicklas. 2023. Privacy-aware Publication of Wi-Fi Sensor Data for Crowd Monitoring and Tourism Analytics. In *Proc. of the 1st ACM SIGSPATIAL Intl. Workshop on Geo-Privacy and Data Utility for Smart Societies*. ACM, Hamburg Germany. doi:10.1145/3615889.3628513
- [2] Leonie Ackermann, Michael Mühlhauser, Alexandru Burdusel, Michael Federlin, Dominik Herrmann, Steffen Holly, Daniela Nicklas, and Daniel Wolpert. 2023. Towards Anonymizing Intermodal Mobility Data for Smart Cities. In *Proc. of the 1st ACM SIGSPATIAL Intl. Workshop on Geo-Privacy and Data Utility for Smart Societies*. ACM, Hamburg Germany. doi:10.1145/3615889.3628514
- [3] Jakob E. Bardram. 2022. From Sensing to Acting—Can Pervasive Computing Change the World? *IEEE Pervasive Computing* 21, 3 (July 2022). doi:10.1109/MPRV.2022.3182489
- [4] Yacine Belal, Sonia Ben Mokhtar, Hamed Haddadi, Jaron Wang, and Afra Mashhadi. 2024. Survey of federated learning models for spatial-temporal mobility applications. *ACM Transactions on Spatial Algorithms and Systems* 10, 3 (2024).
- [5] Debasree Das, Rahat Rafiq, and Daniela Nicklas. 2025. Does One Noise Fit All? Analyzing Utility in Transportation Mode-Based Anonymization. In *2025 IEEE Intl. Conf. on Smart Computing (SMARTCOMP)*. IEEE.
- [6] Ziquan Fang, Dongen Wu, Lu Pan, Lu Chen, and Yunjun Gao. 2022. When Transfer Learning Meets Cross-City Urban Flow Prediction: Spatio-Temporal Adaptation Matters. In *IJCAI*, Vol. 22. 2030–2036.
- [7] Mila Gascó. 2016. What Makes a City Smart? Lessons from Barcelona. In *49th Hawaii Intl. Conf. on System Sciences, HICSS 2016, Koloa, HI, USA, January 5–8, 2016*, Tung X. Bui and Ralph H. Sprague Jr (Eds.). IEEE Computer Society. doi:10.1109/HICSS.2016.373
- [8] Yilun Jin, Kai Chen, and Qiang Yang. 2022. Selective Cross-City Transfer Learning for Traffic Prediction via Source City Region Re-Weighting. In *Proc. of the 28th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD ’22)*. 731–741. doi:10.1145/3534678.3539250
- [9] Ivan Kralj, Lodovico Giarretta, Gordan Ježić, Ivana Podnar Žarko, and Šarūnas Girdzijauskas. 2024. Semi-decentralized Training of Spatio-Temporal Graph Neural Networks for Traffic Prediction. *arXiv preprint arXiv:2412.03188* (2024). <https://arxiv.org/abs/2412.03188>
- [10] Wenzhu Li and Shuang Wang. 2022. Federated meta-learning for spatial-temporal prediction. *Neural Computing and Applications* 34, 13 (2022).
- [11] Christian Oberli, Miguel Torres-Torriti, and Dan Landau. 2010. Performance Evaluation of UHF RFID Technologies for Real-Time Passenger Recognition in Intelligent Public Transportation Systems. *IEEE Transactions on Intelligent Transportation Systems* 11, 3 (Sept. 2010). doi:10.1109/ITITS.2010.2048429
- [12] Masakazu Ohno, Riki Ukyo, Tatsuya Amano, Hamada Rizk, and Hirozumi Yamaguchi. 2024. Privacy-preserving pedestrian tracking with path image inpainting and 3D point cloud features. *Pervasive and Mobile Computing* 100 (2024).
- [13] Ren Ozeki, Haruki Yonekura, Aidana Baimbetova, Hamada Rizk, and Hirozumi Yamaguchi. 2023. One Model Fits All: Cross-Region Taxi-Demand Forecasting. In *Proc. of the 31st ACM Intl. Conf. on Advances in Geographic Information Systems (SIGSPATIAL ’23)*. 1–4.
- [14] Ren Ozeki, Haruki Yonekura, Hamada Rizk, and Hirozumi Yamaguchi. 2024. Privacy Preserved Taxi Demand Prediction System for Distributed Data. In *Proc. of the 32nd ACM Intl. Conf. on Advances in Geographic Information Systems (SIGSPATIAL ’24)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3678717.3691234
- [15] Tim Rutermaun, Aboubakr Benabbas, and Daniela Nicklas. 2019. Know Thy Quality: Assessment of Device Detection by WiFi Signals. In *IEEE Intl. Conf. on Pervasive Computing and Communications Workshops, PerCom Workshops 2019, Kyoto, Japan, March 11–15, 2019*. IEEE. doi:10.1109/PERCOMW.2019.8730828
- [16] Rijurekha Sen, Youngki Lee, Kasthuri Jayarajah, Archan Misra, and Rajesh Krishna Balan. 2014. GruMon: fast and accurate group monitoring for heterogeneous urban spaces. In *Proc. of the 12th ACM Conf. on Embedded Network Sensor Systems (SenSys ’14)*. doi:10.1145/2668332.2668340
- [17] Chong Shang, Haizhou Ai, and Bo Bai. 2016. End-to-end crowd counting via joint learning local and global count. In *2016 IEEE Intl. Conf. on Image Processing (ICIP)*. 1215–1219. doi:10.1109/ICIP.2016.7532551 ISSN: 2381-8549.
- [18] Venkatesh Bala Subburaman, Adrien Descamps, and Cyril Carincotte. 2012. Counting People in the Crowd Using a Generic Head Detector. In *2012 IEEE Ninth Intl. Conf. on Advanced Video and Signal-Based Surveillance*. doi:10.1109/AVSS.2012.87
- [19] Fukuharu Tanaka, Tatsuya Amano, Akira Uchiyama, Akihito Hiromori, Yusuke Nakamura, and Hirozumi Yamaguchi. 2024. Policy Optimization for Pedestrian Traffic Management by Surrogation of Simulation Models. In *2024 IEEE 21st Intl. Conf. on Mobile Ad-Hoc and Smart Systems (MASS)*.
- [20] Masashi Uegaki, Amano Tatsuya, and Hirozumi Yamaguchi. 2024. Simulating Urban Pedestrian Flows by Fusing Wide-Area Location Data and Spot Pedestrian Counts. In *Proc. of the 21st EAI Intl. Conf. on Mobile and Ubiquitous Systems: Computing, Networking and Services*.
- [21] Riki Ukyo, Tatsuya Amano, Hamada Rizk, Hirozumi Yamaguchi, and Takao Moriya. 2025. Robust Pedestrian Tracking With Severe Occlusions in Public Spaces Using 3D Point Clouds. *IEEE Transactions on Intelligent Transportation Systems* 26, 6 (2025).
- [22] Leye Wang, Xu Geng, Xiaojuan Ma, Feng Liu, and Qiang Yang. 2019. Cross-City Transfer Learning for Deep Spatio-Temporal Prediction. In *Proc. of the Twenty-Eighth Intl. Joint Conf. on Artificial Intelligence (IJCAI ’19)*. doi:10.24963/ijcai.2019/262
- [23] Jens Weppner and Paul Lukowicz. 2013. Bluetooth based collaborative crowd density estimation with mobile phones. In *2013 IEEE Intl. Conf. on Pervasive Computing and Communications (PerCom)*. doi:10.1109/PerCom.2013.6526732
- [24] Banglie Yang, Runze Li, Yijing Wang, Sha Xiang, Shuo Zhu, Cheng Dai, Shengxin Dai, and Bing Guo. 2025. Cross-city transfer learning for traffic forecasting via incremental distribution rectification. *Knowledge-Based Systems* 315 (2025).
- [25] Muhammad Yaqub, Shahzad Ahmad, Malik Abdul Manan, and Imran Shabir Chuhan. 2024. Predicting Traffic Flow with Federated Learning and Graph Neural with Asynchronous Computations Network (FLAGCN). *arXiv preprint arXiv:2401.02723* (2024). <https://arxiv.org/abs/2401.02723>
- [26] Qi Zhang and Antoni B. Chan. 2019. Wide-Area Crowd Counting via Ground-Plane Density Maps and Multi-View Fusion CNNs. In *2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/CVPR.2019.00849 ISSN: 2575-7075.
- [27] Yixin Zhang, Tianyu Zhao, and Salma Elmalaki. 2024. Towards Fairness-aware Crowd Management System and Surge Prevention in Smart Cities. In *2024 IEEE Workshop on Design Automation for CPS and IoT (DESTION)*.