Ackermann, Leonie; Akcabay, Samet; Benabbas, Aboubakr; Khalil, Reem Eslam ; Nicklas, Daniela

Enhancing Data Quality and Collaboration in Participatory Climate Data Crowdsensing

Primary publication

2024 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops) - New York : IEEE, 2024 - DOI: 10.1109/percomworkshops59983.2024.10502897.

Enhancing Data Quality and Collaboration in Participatory Climate Data Crowdsensing

***Abstract*—Based on a citizen-initiated climate data crowdsensing campaign, we showcase mechanisms for ensuring data quality and fostering collaborative curation of the collected information. Specifically, we delve into the challenges and potentials inherent in citizen-driven data collection, leveraging the Netatmo Smart Home Weather Station. Our study addresses the engagement of citizens in the scientific process, investigates innovative detection mechanisms for enhancing data quality in crowdsensed climate data, and introduces robust metrics for assessing data quality. Furthermore, we introduce a prototype application designed to empower citizens through active participation in data validation and curation processes, thereby elevating the trustworthiness and utility of the collected data. These solutions actively contribute to the ongoing design and evolution of climate data crowdsensing, harnessing the potential of the urban knowledge society. In this way, they actively support the participation of the city community in overcoming social challenges associated with climate change adaptation.**

***Index Terms*—data quality, citizen participation, crowdsensing, climate data**

I. Introduction

In the face of climate change, cities worldwide are grappling with the imperative to adapt swiftly and effectively to shifting environmental conditions. Urban areas, often characterized by impervious surfaces and extensive infrastructure, find themselves particularly vulnerable to the multifaceted impacts of climate change, ranging from intensified heatwaves to increased frequency of extreme weather events. In 2022, a citizens' association started to set up a climate monitoring network in PseudoTown[1] to gain a better understanding of the impact of climate change on the urban microclimate. Using ten smart weather stations from Netatmo[2], the network measures the temperature and humidity at various locations within the city. The data can be viewed in real time on an interactive map in a web application[3] designed for this purpose. Over fifty other Netatmo stations are in operation in PseudoTown. Each of these stations' data can be accessed in real-time from the Netatmo website[4].

With the climate monitoring network, the citizens' association aims to gain a comprehensive understanding of the urban heat island effect [1] in PseudoTown and to influence decision makers, such as politicians, urban planners, and authorities, to create a more sustainable city. However, errors and anomalies in the readings of the smart weather stations are a setback in achieving this goal: Due to exogenous influences on the stations, outliers occur in the readouts, negatively affecting the sensor data. Hence, decision-makers might not trust the data produced by citizens and refrain from taking the necessary actions.

In this scenario, we can observe different dimensions of trust:

- Trust in the data: Depending on the quality of the sensor system and the installation, a single sensor might provide more or less accurate measurements. Decision makers tend to trust official and standardized measurement stations since the produced data better reflects the observed phenomena.
- Trust in the information: Accuracy is only one of many data quality dimensions. When it comes to distributed phenomena like micro-climate or air quality, spatial coverage gains importance. We can only derive higher-level information from single data points if we address all relevant quality dimensions, which depend on the application domain.
- Trust in the decision: Decision makers interpret the information and take action. A lack of information quality can lead to bad decisions; if, for example, the citizens get the impression that the city uses only selected locations for their measurements, they might lose trust in these actions.

Hence, climate monitoring aims to improve the quality of information by densifying the data collection and ensuring that relevant locations (e.g., local heat islands) are reflected in the data. However, the challenge of controlling data quality remains. The members of the citizens' association face the time-consuming task of analyzing the stations' outputs, applying quality control tools, and identifying missing data.

As illustrated in Figure 1, we follow a layered approach to provide trustworthy information from raw sensor data. As a first step, we apply automated data quality control methods to identify potential problems in the data. These could be quality patterns like those proposed in [2]; in this paper, we adapted the CrowdQC+ method [3] to our use case and evaluated the outcomes with domain experts. On top of that, we work on tools to support a participatory data validation approach. A first prototype supports domain experts and involved citizens to inspect potential data quality issues and to make informed decisions on whether to include the flagged data in analyses

[1]Medium-sized German town with a historic city center. Place name pseudonymized to ensure the anonymity of submission. Real name is used for Camera Ready version.

[2]https://www.netatmo.com/de-de/smart-weather-station (accessed 2023-11-29)

[3]https://weathermap.netatmo.com/ (accessed 2023-11-29)

[4]https://dev.netatmo.com/apidocumentation/weather (accessed 2023-11-29)

(because it represents an important real-world phenomenon) or to exclude it (because it might weaken the information quality).

The paper is structured as follows: Section II introduces the opportunities and challenges of crowd-sensed climate data, the role of Netatmo Weather Stations, and the data quality control framework of CrowdQC+. Section III explores detection mechanisms for data quality in crowd-sensed climate data, presents metrics for evaluating data quality, and addresses challenges unique to climate data. Section IV focuses on the active involvement of citizens in data validation and curation processes, emphasizing the collaborative effort to enhance the trustworthiness of the collected data, while Section V summarizes the main outcomes and provides an outlook on future work.

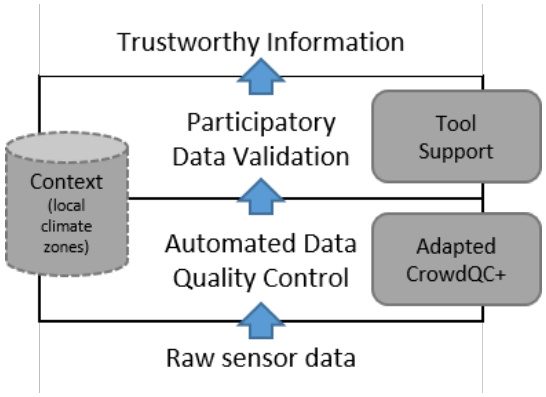


Fig. 1. Layered Approach for Trustworthy Information

II. Background and Related Work

Ray et al. [4] refer to mobile crowdsensing as the wide variety of sensing models in which individuals collectively share data and extract information to measure and map phenomena of common interest using smart mobile devices. According to Ganti et al. [5], the measurement can be divided into two different categories: First, personal sensing, in which participants measure personal data because of their own interest (health monitoring, personal carbon footprint). Second is community sensing, in which real phenomena cannot be measured by a single person but only by groups of individuals. This sensing category can also be divided into two different types: participatory sensing [6], which represents the focus of this work, and opportunistic sensing [7]. The first type requires active participation supported by personal motivation (e.g., via providing captured images, the input of information, etc.), whereas with the second type, measurements and data are collected independently (e.g., temperature measurements on smartphones without active input of users).

A. Crowdsensing of Environment Data

To contextualize PseudoTown's climate monitoring network with other initiatives' crowdsensing environment data, we examine two projects: *Crowdsensing für Bodensee Online* and the *Budburst* application.

Crowdsensing für Bodensee Online is a project initiated by ISB AG, the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation and Engineering Company Prof. Kobus und Partner GmbH in February 2020 to measure the water temperature of Lake Constance with the help of boat owners using mobile sensors and maps [8]. The project's crowdsensing approach is considered suitable for improving and refining an existing (data) structure, as high data quality and density can be achieved in a small space, provided that sufficient quality assurance and data quantity are available Due to the underlying characteristics, *Crowdsensing für Bodensee Online* can be classified as opportunistic sensing, as the sensors only need to be attached to the boats, and no active participation of the boat owners is necessary.

The *Budburst*[5] application encourages users to identify plant species using their own cell phone and to record local plant (species) together continuously. This encourages users to monitor the state of the local ecosystem to draw conclusions about climate change. Photographing plants with the app stimulates a discourse on the ecosystem's health, which the users and the resulting community lead. The *Budburst* app is an example of participatory sensing, as users are actively involved in the data collection.

Bodensee Online and *Budburst* offer users a visualization of the sensor outputs, thus enabling them to draw their own conclusions from the data. For *Bodensee Online*, for example, clicking on the interactive map for the respective position, graphs are displayed showing the course of the water temperature in °C in relation to the current in m/s[6]. The display can be manipulated at will by time and water depth filtering.

The *Budburst* application, on the other hand, shows the course of the data collected for a location: Depending on which location has how many entries for a particular plant, users can click on an individual point on the map to view the history and the entries made by other users for that point. Users can then see the extent to which the observed plant has changed over time through status diagrams. Both applications allow for analyzing the recorded data by displaying graphs and diagrams.

B. Data Quality in Crowdsensed Data

Crowdsensing enables more diverse data to be collected in larger quantities in a decentralized manner. However, the larger amount of collected data results in many redundancies and duplicates, which leads to a need for quality control of the results [4]. Data in crowdsensing projects is usually collected by large networks of sensors that are not highly calibrated and accurate. These sensors need close quality control for their data. To do so, such large sensor groups or networks need to be presented as such using a model that makes it easy to find sensors and relate them to their neighbors for data quality control. Benabbas et al. [9] provided a semantic model using the Semantic Sensor Network ontology [10] to

[5]https://budburst.org/ (accessed 2023-11-30)

[6]https://www.lubw.baden-wuerttemberg.de/wasser/bodenseeonline (accessed 2023-11-30)

gather information about sensors from a citizen science project called "Sensor.Community"[7]. The sensor models created are used to create data processing flows to perform quality-aware data processing on the sensors stored in the model [11]. Data quality control entails defining the different dimensions that quantify its quality. Batini et al. [12] give a thorough overview of the different definitions of quality through different dimensions and metrics. Teh et al. compiled in their survey [13] a comprehensive list of research contributions dedicated to sensor data quality control.

C. Crowdsensing with Netatmo Weather Stations

Multiple approaches exist for collecting temporal data, with the primary challenge lying in avoiding biases introduced by different manufacturers, each with its own set of biases and error characteristics [14]. Netatmo stations are distributed worldwide; in PseudoTown alone, there are approximately 60 stations, and their numbers continue to increase. The Netatmo company develops and distributes these stations to citizens interested in atmospheric conditions monitoring outdoors and indoors [14].

Netatmo sensors are cost-efficient, offering WiFi connectivity for data transfer. They automatically upload their data to the Netatmo server and provide real-time data visualization through application software [14]. To access the data through an API request, users can easily create an account on the Netatmo website and generate an access token for the desired stations. Furthermore, users can obtain the station's metadata for the region of interest through an API request by providing the longitudes and latitudes of the region.

A study conducted in Berlin using Netatmo stations and reference monitoring stations revealed a deviation between Netatmo stations and the referenced stations used in the study [15]. Despite this deviation, the study demonstrated great potential in Netatmo stations for long-term temperature monitoring. Applying the right data quality techniques could enhance their usefulness for climate research.

D. CrowdQC+ as Temperature Data Quality Tool

CrowdQC+ is built on top of a quality control package called CrowdQc [3]. It inherits all the quality control levels of CrowdQc, along with additional quality control levels and functions. CrowdQC+ has five main quality check levels:

- Level m1: Metadata check to ensure that no two stations share the same longitude and latitude. Otherwise, all stations with identical longitude and latitude will not be considered in subsequent levels.
- Level m2: Distribution check to ensure all station observations' timestamps have the required gap. Additionally, it involves removing observations with radiative errors.
- Level m3: Data validity check to ensure, by default, that each station has 80% valid observations. Otherwise, the station will not be considered at further levels.
- Level m4: Temporal correlation check, examining the temporal relation between each station and the median of all stations for a specified period. This level also aims to remove radiation errors or indoor stations.
- Level m5: Spatial correlation check to assess the correlation between each station's observations and those of other stations in the same neighborhood. Stations' neighbors are identified within a default radius of 3000 meters. Stations with fewer than five neighbors are considered isolated and, by default, are excluded. This level removes stations that survived other levels but still have faulty values, primarily single unrealistic high values due to radiative errors.

A study in Amsterdam and Toulouse used Netatmo stations quality controlled by CrowdQC and CrowdQC+. Professionally operated weather stations were also quality-controlled and utilized to evaluate both QC packages. The study revealed that CrowdQC+ exhibited fewer deviations than CrowdQC, with more noticeable improvements on warmer days. However, CrowdQC+ had lower data availability, which is understandable given its incorporation of additional filters. The findings indicated that CrowdQC+ effectively removes erroneous data, even though some unfiltered faulty observations may still persist.

Nevertheless, we believe that the spatial correlation at Level m5 has a significant weakness, as local climate zones are not considered. We will outline this concept in the next subsection.

E. Local Climate Zones

The structure of a city influences temperature, leading to a phenomenon known as the urban heat island effect. This effect occurs when an urban area[8] experiences significantly higher temperatures than its surrounding rural areas[9] [1]. This phenomenon arises due to the artificial surfaces prevalent in urban areas, especially during the day. The heat absorbed from the morning sun is retained in urban fabric materials such as concrete, brick, and tiles, affecting each monitoring station based on its local climate zone. This may include or exclude observations in Level 5 of CrowdQC+ based on potentially incorrect station neighbors.

A crucial aspect of understanding the urban heat island involves classifying urban and rural areas, a task facilitated by delineating different types of local climate zones. Oke's scheme divides the city's terrain into seven homogeneous regions called local climate zones [16]. Each zone is classified based on the buildings' height and intensity of land cover types.

III. Detection Mechanisms for Data Quality in Crowdsensed Climate Data

In the following, two modifications for the spatial correlation in level m5 of CrowdQC+ are presented, and their accuracy is evaluated compared to the original CrowdQC+.

[7]https://sensor.community, last access: 05.11.2023

[8]Urban areas: areas with high population density and a well-developed built environment

[9]Rural areas: areas with low-density populations and small settlements

A. Approach 1: Assigning Sensor Locations to Local Climate Zones

Neighbors are defined based on their local climate zones instead of a constant distance approach. Oke's scheme [16] was used to classify the local climate zones in PseudoTown. Using Google Earth, polygons were created for each zone with the help of a domain expert. Subsequently, the stations were added based on the latitude and longitude coordinates and then assigned to the correct zone.

B. Approach 2: K-means clustering and elbow curve

A certain number of clusters is defined, and stations are partitioned to these clusters using k-mean clustering. Data clustering is a data mining technique for grouping data with similar characteristics. The k-means algorithm is a popular method for clustering data, giving the number of clusters. Finding the optimal number of clusters is a trial and error process as the performance of the clustering algorithm may be affected by the chosen K value [17]. The optimal number of clusters needed for partitioning stations is determined using the elbow curve method (Approach 2a). Additionally, Approach 2b was implemented, which assigns the same number of clusters as the number of local climate zones in the region of interest.

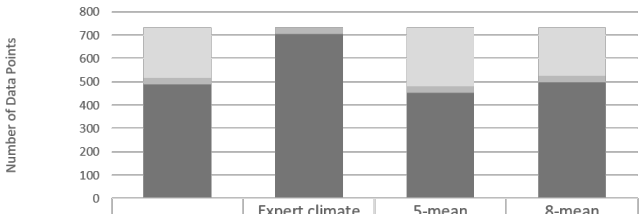
C. Evaluation

The data for PseudoTown in March and August was collected from sixty different Netatmo stations. The data underwent quality control through three distinct approaches: CrowdQC+, Approach 1, and Approach 2 (2a, with five clusters determined by elbow curve, and 2b, eight clusters aligning with the number of local climate zones). The outputs are divided as follows: The number of,

- True values that survived all the data quality tests
- False values because of higher temperature than other stations during sun time
- False values because of lower temperature than other stations during sun time
- False values because of higher temperature than other stations during nighttime
- False values because of lower temperature than other stations during nighttime
- False values because of failure in the temporal correlation test
- False values because of failure in the spatial correlation test

Figure 2 is an example of the output for Ottostraße presented to a domain expert in August. The expert was not informed which result is attributable to which approach, and he should assess whether the classification as a correct or incorrect value is plausible from his point of view. The expert considered the results of approach 1 to be the most conclusive.

As shown in Figure 2, false values using the local climate zone approach have the fewest inaccuracies. Outputs from other methods, namely CrowdQC+, five means cluster, and eight means cluster, exhibit more false values due to a failure in spatial correlation. This indicates that these methods incorrectly categorized the Ottostraße station into the wrong cluster.



	CrowdQC+	Expert climate zoning	5-mean clustering	8-mean clustering
Spat_Corr	217	0	253	208
Temp_Corr	0	0	0	0
Night_Low	26	26	26	26
Night_High	0	0	0	0
Day_Low	1	1	1	1
Day_High	2	2	2	2
True data	487	704	451	496

Fig. 2. Distribution of Detected Errors (Ottostraße, August 23) - Total: 733

The experiment showed that, while CrowdQC+ is a robust data quality tool, its efficacy could be further improved by incorporating considerations for local climate zones in level m5 clustering. The developed mechanisms provide a solid foundation for the curation tool, which supports professionals and committed citizens in reviewing potential data quality issues and making informed decisions about including the flagged data in analyses. We will discuss the prototype in the next section.

IV. Collaborative Validation of Crowdsensing Data

To enable citizens and domain experts to validate and curate data, we implemented a prototype that imports the quality-controlled data from CrowdQC+ and visualizes them in a web application for further analysis. The requirements, implementation, and preliminary results are presented in the following subsections.

A. Requirements

The requirements were collected iteratively in regular meetings with the citizens' association and a domain expert. User stories are outlined below to illustrate the requirements using a specific example. The user stories are based on the statements of the stakeholders during the regular meetings; their needs, wishes, and suggestions are intended to provide a concrete vision for various people (or groups), which is supported by the development of an interactive tool.

1) **Citizen** I would like to be able to view the temperature and humidity at the various locations in real time to recognize the effects of climate change on the city.
2) **Member of the citizens' association** I would like to view and analyze the temperature and humidity at the various locations in the city center in real time to identify

the effects of impervious surfaces in PseudoTown's city center.

3) **Member of the German Cyclists' Federation (ADFC)** I would like to monitor the temperature and air humidity at the various locations to investigate the effects of car traffic on PseudoTown's climate.
4) **Meteorologist** I would like to ask citizens who operate weather stations what exogenous effects act on the sensors when measurements seem unusual to me so that I can be sure whether it is a genuine microclimatic deviation or a measurement error.

B. Implementation

The front end of the application is implemented with the Python-Flask[10] web framework. Flask uses Jinja2[11] as the template engine, with which HTML and JavaScript code and CSS stylesheets can be integrated. The open source software Grafana[12] is used to visualize and interact with the data. Grafana can retrieve data from a database, visualize it individually, and enter it into dashboards.

Eight dashboards were implemented in the collection of PseudoTown's weather stations, with one dashboard consolidating the other seven weather stations to enable a comparison between them. The focus here is on the visualization of the temperature at a given time for each weather station. Temperature ranges can be highlighted in color using threshold values that can be defined in the dashboards. The aspect of collaborative data validation is supported natively by Grafana through the ability to mark a point on the graph and add an annotation to this marker. Other users of the tool can then view these annotations.

The front end offers users four pages:

- **Home** for the application's start page. All weather stations and the defined Local Climate Zones are displayed on an OpenStreetMap map. Users can access a chat window to exchange information about the data on the map with other users.
- **Sensor inspector** for analyzing the data from the PseudoTown weather stations. On this page, the data from Grafana is integrated using iFrames and restricted by filter options (for weather stations, date, and time) using dropdown menus.
- **Feedback** for the opportunity to provide feedback on the application. Here, users can leave a message and assign a category to it (error message, change suggestion, or new feature).
- **User** for user administration. Here, it is possible to register and log in. After logging in, users can view their profiles, chats, and feedback. Favorite weather stations can also be managed in this section.

The application's back-end is implemented with PostgreSQL[13] as a relational database management system. PostgreSQL was chosen as the database system because it is an open-source software that can be easily integrated into Grafana. Sensor data from March 1, 2023, to August 10, 2023, was imported into the database from CSV files. An overview of the application architecture is provided in Figure 3.

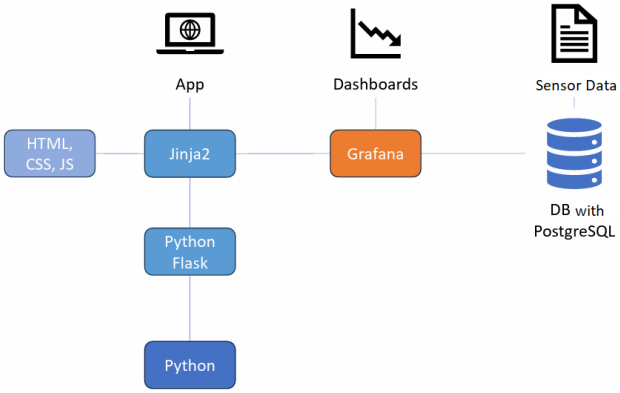


Fig. 3. Application Architecture

C. Results and Limitations

The prototype development process and its limitations are discussed below. The requirements set by the stakeholders were largely fulfilled, and a first prototype was successfully developed. Implementing the aspect of data validation and curation in the application was challenging. Including functionalities allowing users to comment on the existing data is a sensible but complex solution to implement. Simple input methods attached to specific data sets (and ideally linked to a database) can be used to build user data validation into applications. The development of the prototype benefited from an iterative planning process and a thorough validation and prioritization of the components. The collaboration with the stakeholders has been consistently positive.

At the same time, it must be emphasized that the concept, the established requirements, and the application's implementation were tailored to specific stakeholders' needs. Other stakeholders may have different requirements and needs, so the components mentioned should be adapted depending on the context.

V. Conclusion

In conclusion, this paper sheds light on the pivotal role of collaboration between citizens and domain experts in shaping the design and evolution of crowdsensing campaigns, particularly in the context of climate data collection. Our insights contribute to harnessing the potential of urban knowledge societies and underscore the significance of robust data quality metrics and collaborative curation mechanisms for trustworthy data and information. Our exploration of citizen-centric approaches for trustworthy crowdsensing emphasizes the empowerment of communities in shaping sustainable transformation.

In our future work, we would like to pursue the implications of our findings in several key areas. Automating outlier

[10]https://flask.palletsprojects.com/ (accessed 2023-11-30)
[11]https://jinja.palletsprojects.com/ (accessed 2023-11-30)
[12]https://grafana.com/ (accessed 2023-11-30)
[13]https://www.postgresql.org/ (accessed 2023-11-30)

detection by leveraging machine learning models capable of understanding stations' output patterns over time would be beneficial for predicting typical variations, providing valuable input for CrowdQC+. Moreover, integrating stations' patterns with clustering approaches or automated zone-setting programs holds promise for streamlining data quality evaluation, potentially alleviating the burden on domain experts.

Implementing the curation tool introduced in this paper should undergo rigorous testing by stakeholders in future work. Feedback on usability and effectiveness is crucial for refining the tool's functionality. Comparative analyses between previous manual analyses of citizens' weather station outputs and analyses using the curation tool will provide insights into the effectiveness of the collaborative data validation and the curation approach. We will further explore the dynamics of the urban knowledge society's contributions to the public interest-oriented use of information, moving from the role of users to active data producers.

REFERENCES

[1] J. Voogt, "How researchers measure urban heat islands," in *United States Environmental Protection Agency (EPA), State and Local Climate and Energy Program, Heat Island Effect, Urban Heat Island Webcasts and Conference Calls*, 2007.

[2] A. Benabbas, S. Steuer, and D. Nicklas, "Towards Adaptive Sensor Data Quality Improvement based on Context Models," in *2020 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2020, Austin, TX, USA, March 23-27, 2020*. IEEE, 2020, pp. 1–6. [Online]. Available: https://doi.org/10.1109/PerComWorkshops48775.2020.9156173

[3] D. Fenner, B. Bechtel, M. Demuzere, J. Kittner, and F. Meier, "Crowdqc+—a quality-control for crowdsourced air-temperature observations enabling world-wide urban climate applications," *Frontiers in Environmental Science*, vol. 9, p. 553, 2021.

[4] A. Ray, C. Chowdhury, S. Bhattacharya, and S. Roy, "A survey of mobile crowdsensing and crowdsourcing strategies for smart mobile device users," *CCF Transactions on Pervasive Computing and Interaction*, vol. 5, no. 1, pp. 98–123, Mar. 2023. [Online]. Available: https://doi.org/10.1007/s42486-022-00110-9

[5] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 32–39, 2011. [Online]. Available: https://doi.org/10.1109/MCOM.2011.6069707

[6] J. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory sensing," in *Workshop on World-Sensor-Web (WSW'06): Mobile Device Centric Sensor Networks and Applications*, 2006, pp. 117–134.

[7] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, "A survey of mobile phone sensing," *IEEE Communications magazine*, vol. 48, no. 9, pp. 140–150, 2010.

[8] ISB AG, Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung, and Ingenieurgesellschaft Prof. Kobus und Partner GmbH, "Das Projekt CrowdSensing für Bodensee Online: Erfahrungsbericht der Projektbeteiligten," Ministerium für Umwelt, Klima und Energiewirtschaft Baden-Württemberg, Stuttgart, Tech. Rep., Jul. 2021. [Online]. Available: https://um.baden-wuerttemberg.de/fileadmin/redaktion/m-um/intern/Dateien/Dokumente/3_Umwelt/Nachhaltigkeit_und_Umweltbildung/Digitalisierung/CrowdSensing-fuer-Bodensee-Online-Zusammenfassung.pdf

[9] B. Aboubakr, H. Hannes, and N. Daniela, "Semi-Automatic Ontology Population for Online Quality Assessment of Particulate Matter Sensors," in *Workshop Proceedings of the 14th International Conference on Intelligent Environments, Rome, Italy, 25-28 June 2018*, ser. Ambient Intelligence and Smart Environments, Rome, 2018, pp. 119–128. [Online]. Available: http://ebooks.iospress.nl/volumearticle/49355

[10] Michael Compton et al., "The SSN ontology of the W3C semantic sensor network incubator group," *Web Semantics: Science, Services and Agents on the World Wide Web*, 2012.

[11] A. Benabbas, M. Geißelbrecht, G. M. Nikol, L. Mahr, D. Nähr, S. Steuer, G. Wiesemann, T. Müller, D. Nicklas, and T. Wieland, "Measure particulate matter by yourself: data-quality monitoring in a citizen science project," *J. Sens. Sens. Syst.*, vol. 8, no. 2, pp. 317–328, Nov. 2019, publisher: Copernicus Publications. [Online]. Available: https://jsss.copernicus.org/articles/8/317/2019/

[12] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, "Methodologies for data quality assessment and improvement," *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–52, 2009.

[13] H. Y. Teh, A. W. Kempa-Liehr, and K. I.-K. Wang, "Sensor data quality: A systematic review," *Journal of Big Data*, vol. 7, no. 1, pp. 1–49, 2020.

[14] J. Coney, B. Pickering, D. Dufton, M. Lukach, B. Brooks, and R. R. Neely III, "How useful are crowdsourced air temperature observations? an assessment of netatmo stations and quality control schemes over the united kingdom," *Meteorological Applications*, vol. 29, no. 3, p. e2075, 2022.

[15] F. Meier, D. Fenner, T. Grassmann, M. Otto, and D. Scherer, "Crowdsourcing air temperature from citizen weather stations for urban climate research," *Urban Climate*, vol. 19, pp. 170–191, 2017.

[16] I. D. Stewart and T. R. Oke, "Local climate zones for urban temperature studies," *Bulletin of the American Meteorological Society*, vol. 93, no. 12, pp. 1879–1900, 2012.

[17] D. T. Pham, S. S. Dimov, and C. D. Nguyen, "Selection of k in k-means clustering," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 219, no. 1, pp. 103–119, 2005.